Higher-Order Networks in Complex Systems

Temporality and Interconnectivity

DISS. ETH Nr. 23457

Higher-Order Networks in Complex Systems: Temporality and Interconnectivity

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZÜRICH
(Dr. sc. ETH Zürich)

presented by
NICOLAS WIDER

MSc ETH Mathematics, ETH Zürich, SWITZERLAND

born on APRIL 6, 1987

citizen of ZÜRICH, SWITZERLAND

accepted on the recommendation of
PROF. DR. DR. FRANK SCHWEITZER
PROF. DR. VITO LATORA

2016

ETH zürich

Acknowledgments

This thesis represents the fruitful conclusion of a three years long journey. The path was scattered with challenging tasks and detours exploring the variety of interdisciplinary research. Thankfully, I was not alone on this path and could draw on the support of several people that guided me along the way.

First and foremost I want to thank my advisor Frank Schweitzer for his continuous help and support throughout the thesis. His immense experienced helped me a lot to really understand what it means to be a scientist and how to shape the scientific profile. The invaluable comments he provided on any matter be it about scientific or general skills will beneficial on my future paths. The Chair of System design offered a pleasant and encouraging environment for which I am very thankful.

Special thanks go to my closest collaborators Antonios Garas and Ingo Scholtes. They accompanied me during the whole process and provided helpful advices. Their inspiration and motivation helped me to overcome any kind of obstacles and to always appreciate the qualities of scientific work. All of the research presented in this thesis was conceived together with them and I was very privileged to work with such experienced scientists.

I am also thankful to all the other, current and former, members of the Chair of Systems Design that I had the pleasure to work with. They provided a friendly and comfortable atmosphere, be it general discussions that broadened the usual scientific scope or in giving helpful support and feedback on particular aspects of my work. The interdisciplinary environment allows to learn new things every day and to perceive the own research from different perspectives.

On a personal level, I am grateful to my parents and family who supported me in any manner and made sure that I did not have to worry about any issue of everyday life and thus could focus on my research. Further, I thank my close friends and colleagues which continuously reminded me to also enjoy the non-science related aspects of life.

Contents

Abstract

A particular approach to study complex systems in nature, society and technology is network theory. In this perspective actors in the system are represented as nodes and interactions or connections among the actors as links. Although the usual framework is often sufficient to study complex systems, it neglects additional properties that are not easily captured. The focus of this thesis is the expansion of classical approaches in network theory that allow to overcome some of the limitations imposed by oversimplifications. We provide *higher-order* models that take into account additional information of complex systems, in particular the aspects of temporality and interconnectivity.

In the first part we study *temporal networks* in which the interactions or connections that occur in the systems are tagged with a time-stamp. This allows to study different aspect of temporal interaction sequences, such as the duration of interactions or inter-event times. In this work we focus on the *ordering* of links, rather than the *timing*. To study temporal ordering we develop a framework of *higher-order time-aggregated networks*, a generalization of the commonly applied static, time-aggregated representation of temporal networks. This approach explicitly captures temporal path statistics and therefore is able to preserves causality structures. With this tool at hand we discuss and analyze different application scenarios that reveal the need of higher-order models and that highlight the power of the proposed approach. More precisely, we study path-based centralities and dynamical processes in respect of temporal ordering.

In the second part we investigate coupled systems that are *interconnected*, meaning that a system is not only influenced by the connections among its own nodes but also by connections to other systems. Recent research has addressed this issue with so-called *multi-layer* networks. First, we investigate the *lack information* one is often confronted with when analyzing real-world systems. The detailed multi-layer topology can be missing and only aggregated statistics may be available. We study dynamical processes in different cases where we rely on limited information of the link topology. Next, we focus on a system of scholarly publications that combines different types of links and nodes that are connected to each other. Leveraging on the multi-layer topology we investigate how co-authorships are correlated with citations and how this affects citation based ranking schemes.

Concluding, this thesis provides a deeper understanding of complex systems and goes beyond the commonly used static and decoupled network approach. We not only show the need of an extended perspective but also provide models that overcome the aforementioned limitations. Validating our models on empirical data sets we contribute new insights and techniques that can be used in a broad variety of applications.

Kurzfassung

Netzwerk Theorie ist eine Methode um komplexe Systeme in der Natur, Gesellschaft oder Technologie zu erforschen. Diese Theorie stellt Akteure eines System als Knoten dar und Interaktionen oder Verbindungen zwischen ihnen als Links. Obwohl das übliche Framework oft ausreicht um komplexe Systems zu untersuchen, vernachlässigt es weitere Eigenschaften, die nicht einfach zu erfassen sind. Der Fokus dieser Arbeit ist die Erweiterung der klassischen Netzwerk Theroy um Einschränkungen von Vereinfachung zu überwinden. Wir präsentieren neue Modelle höherer Ordnung, die auch zusätzliche Informationen von komplexen Systemen berücksichtigen, insbesonders Temporalität und Interkonnektivität.

Im ersten Teil untersuchen wir *temporale Netzwerke*, in welchen Interaktionen eines System mit einem Zeitstempel versehen sind. Dies ermöglicht es verschiedene Aspekte, wie die Dauer oder Wartezeiten, von temporalen Interaktionssequenzen zu untersuchen. Anstatt auf das *Timing* fokussieren wir in dieser Arbeit auf die *Ordnung* von Links. Um die temporale Ordnung zu untersuchen entwickeln wir ein zeitaggregiertes Netzwerk höherer Ordnung, eine Verallgemeinerung der üblichen statisch, zeitaggregierten Darstellung eines temporalen Netzwerks. Diese Methode erfasst temporale Pfadstatistiken und bewahrt dadurch die Kausalität. Mit Hilfe dieses Werkzeugs diskutieren und analysieren wir verschiedne Anwedungsszenarien, die die Notwendigkeit von Modellen höherer Ordnung und die Stärke unserer Methode belegen. Im Detail untersuchen wir pfadbasierte Zentralitätsmasse und dynamische Prozesse bezüglich temporaler Ordnung.

Im zweiten Teil erforschen wir vernetzte Systeme, die von den Verbindungen zu anderen Systemen beeinflusst werden. Derzeitige Forschung thematisieren diese Problematik mit sogenannten *multi-layer* Netzwerken. Zuerst untersuchen wir *mangelnde Information*, mit der man oft konfrontiert wird, wenn man reale Systeme analysiert. Die Unkenntnis der genauen multi-layer Topologie kann dazu führen, dass nur aggregierte Statistiken verfügbar sind. Wir untersuchen dynamische Prozesse, in denen wir uns situationsbedingt auf mangelnde Informationen der Link Topologie verlassen müssen. Als nächstes betrachten wir ein System wissenschaftlicher Publikationen, dass verschiedene vernetzte Arten von Links und Knoten kombiniert. Mit Hilfe der multi-layer Topologie untersuchen wir den Zusammenhang von Koautoren und Zitationen, und dessen Einfluss auf zitationsbasierte Ratings.

Diese Arbeit liefert ein vertieftes Verständnis komplexer Systeme und übersteigt übliche Methoden statischer und entkoppelter Netzwerke. Wir demonstrieren nicht nur den Bedarf einer erweiterten Sichtweise sondern offerieren auch Modelle, die die zuvor genannten Einschränkungen überwinden. Durch die Valdierung unser Methode mit empirischen Daten erbringen wir neue Einsichten, die in zahlreichen Anwendungen von Bedeutung sind.

Chapter 1

Introduction

> "Each piece, or part, of the whole of nature is always merely an approximation to the complete truth, or the complete truth so far as we know it. In fact, everything we know is only some kind of approximation, because we know that we do not know all the laws as yet."
>
> Richard Feynman
> *The Feynman Lectures on Physics* (1964)

Science is an expression of humans intrinsic urge to understand nature and the fundamental laws of the world surrounding them. It is driven by acquiring knowledge and trying to figure out the causes of actions. There are several scientific methods to achieve this goal and they vary across scientific disciplines. However, even with the most elaborate methodologies, scientists are not able to fully understand or describe the world as a comprehensive entity. To overcome the difficulty to consider all possible causes one usually focuses on a specific confined *system* [26].

A system is constrained by specific temporal and spatial boundaries and therefore restricts the information that is considered. Usually a system consists of several elements or entities which are observable and therefore can be studied. For example, consider a simple electric circuit consisting of a electrical source such as a battery that is wired to an electric consumer such as a light bulb. This system consists of three basic elements which are the battery, the light bulb and the wires. One could study the current in the wires, the capacity of the battery and the efficiency, resistance or luminosity of the light bulb. Further, one can also study how these properties influence and depend on each other. This simple system can be well understood without knowing much about the surrounding that could affect the circuit. If needed, the influences that come from elements outside

the system, such as an electromagnetic field, could also be imposed as rules, or can be neglected without much loss.

One of the first steps to investigate a particular system is collecting observational data. Analyzing the data allows to explore possible rules and laws that drive the functioning of the system. Understanding systems confronts us with different kind of challenges and difficulties. In the case of the electric circuit described above, assuming some rules and measuring its components is enough to describe what is happening. As another example consider a car that consists of several parts such as the engine, the steering wheel or the spark plugs. Even though from an engineering perspective it may be complicated to build a well performing car, knowing how all of its parts work and how they are assembled is sufficient to assure that the car is functioning and can be operated following some basic driving rules.

This kind of understanding is based on an approach called *reductionism* [42], that assumes that a system can be understood by knowing all of its components. The properties of individual elements of a system and their direct interactions with other elements can be regarded as *micro level*. Meaning, from a microscopic perspective one gets a good understanding of individual elements but one does not have a comprehensive knowledge of the whole system. In contrast to this view one can define the *macro level* that regards the system as a whole. Hence, a macroscopic perspective observes the properties of a system without knowledge about individual elements. For example, observing the behavior of individual people represents the micro level in comparison to observing the society as a whole that represents the macro level. For some systems it is enough to have knowledge about the micro level to also infer properties of the system at the macro level.

However, not all systems can be understood by only analyzing its individual parts or components. Interactions or relations of the system elements can lead to non-trivial effects also implicitly affecting elements that were not part of the immediate interaction. Further, topological structures and hierarchical order between the elements can impose rules that can not be explained based on the micro level. A system exhibiting such properties can be regarded as *complex* and requires more than just collecting data and observations to fully understand it [5, 74].

In other words complex systems are comprised of strongly related elements that lead to emergence of new system qualities. By strongly related we mean that the elements can influence or interact with each other. The term *emergence* relates to the appearance of new system properties that can not be reduced or traced back to the properties of individual elements [75]. To fully understand a complex system the study of both, the micro and macro level, is needed.

Note, that the theory of complex systems refers to a particular perspective or way of understanding a real world system rather then the attempt to classify them. There may be other approaches to deal with a real world system that do not focus on the emergent properties or describe a system in a different manner. However, the complex system approach explicitly focuses on the relationship of micro and macro properties that can not always be understood with other methods.

As an example of a complex system consider the formation of traffic jams. On the micro level each car can be observed according to its speed and direction which they can change according to their surrounding and specified traffic laws. Hence, each car takes a microscopic perspective and acts in respect to their immediate neighbors, like the car in front, in the back or next to them. However, a individual car driver can not foresee the actions of all of the other drivers and therefore can usually only respond to short term observations. If a particular car brakes it will force the subsequent car to brake as well which further can affect additional cars following behind. On the macro level, where we observe the car flow or road capacities, this leads to a traffic jam. It was found that such simple actions can lead to complex behavior affecting the whole system [108] that can not be understood from an analysis of individual cars.

Another example of a complex system is the human brain. Nowadays, biologists and physician know quite well how single neurons work and how they are connected to other neurons. While they are able to infer how neurons interact with each other and how chemical reactions can affect certain parts of the brain the macroscopic outcome in terms of human cognition is still mysterious. The definition and description of consciousness is not only a philosophical challenge but its existence is considered to be an emergent manifestation from complex biological and chemical interactions [160].

Analyzing the various aspects of complex systems got a lot of attention in several fields. Complex systems are present in most scientific disciplines such as physics [113, 142, 146], chemistry [174] and biology [56, 117, 163] but also in social science [173], finance [150] and economy [7, 71, 97]. The complexity of most complex systems emerge from the relations of the system elements. Meaning, that it is not only the property of individual elements that makes as system complex, but rather their relations and interactions with other system elements. Understanding how the entirety of these dyadic relations influence the macro level is therefore often the key to describe and analyze complex systems.

A particular methodology to study complex systems is *network theory*. This approach focuses on the relational links between the system elements. Hence, properties of elements are usually only considered if they result from these links. The links or ties between any two elements and can represent any kind of interaction or connection which makes network

theory a strong tool for a lot of applications. From a mathematical perspective the theory is well described and therefore allows to use precise methods and other mathematical theories to derive conclusions. Over the years the theory of networks evolved and got expanded in several directions. However, the basic framework is usually still constrained by assumptions that limit its direct applicability to real-world systems. Usually, a particular network only represents one type of interaction between the same or similar types of system elements. To overcome this issue one often relies on simplifications such that the well studied tools can be applied in the usual way.

Using the same unaltered methods and models that were already studied for a long time may be tempting and comfortable, however it hinders progress and the emergence of novel tools that may be more feasible in particular situations.

1.1 Higher-order network models in complex systems

Even though network theory proved to be a useful tool to analyze complex systems, it also needs to be adapted to new challenges evolving from applications to real world problems. Novel technologies and the increasing computation power that is available nowadays comes along with new opportunities to analyze complex systems [101]. In earlier times it was sufficient or even desired to restrict the analysis of a certain problem to a well circumscribed and constrained system. The limited computation power and limited applicability of particular methods often required comparably small networks to be able to fully analyze them. Therefore, it was often accepted to acquire approximate results to benefit from the known methods. This way an appropriate and comprehensive representation of a complex system was sacrificed in favor of the analytical precision of a constrained perspective.

In line with the quote by Richard Feynman that introduces this chapter, one is aware that these limitations exist and that they lead to a theory or model that does not agree perfectly with the real world. However, one should cope with this fact and try to improve the theory or methodology by incorporating the laws that we do know. Otherwise, we may be confronted with situations where the simplification of a given problem does not comply at all with the real world. The constrained perspective may produce nice results but still misinterpret the real system under investigation. It is not always straightforward to detect all the issues that come along with a simplified perspective. However, once such a shortcoming is revealed, one should adjust or modify the theory and tool set one relies on.

The increasing availability of vast amount of data allows to represent a given system in much more detail. Incorporating this additional information therefore requires refined

tools and methods. As an example consider the inclusion of the time dimension. Even though a fundamental ingredient of natural processes, it was and still is only indirectly included in network representation of complex systems. Usually, it only serves as boundary condition to constrain the time-window during which observation and data is considered. However, the precise moment when a link between elements of a system were established was only focused on recently [69].

Consider the example of a disease that is spreading in the population by means of human contacts. From the interaction patterns in the population one can infer an interaction network and figure out who is very prone to an infection and who is unlikely to be infected. From this it is possible to approximate how fast and how severe an epidemic could be in the end. However, the inclusion of time-stamps of interactions would improve estimation of the disease spread. How long people stay contagious, once infected, and how long it takes until the disease breaks out are only two factors that influence the spreading. Further, an individual can only be infected by a person that already got the diseases and therefore the ordering of interactions is relevant.

Even though in the previous example it may be obvious that a time-independent network perspective does not capture the system in an accurate way, such simplifications were still used in several studies [6, 83, 112]. In these models the time-evolution of the process applied to the network is considered but not the time-dependency of the links forming the network themselves. Such simplifications are not only applied to time dimensions in networks but also to other properties that are relevant to analyze real-world phenomena. Therefore, more elaborate models are needed to incorporate different interactions and dependencies between the elements of a systems.

To overcome limitation still present in current network approaches, in this thesis we take up the challenge to provide novel higher-order perspectives to complex systems. The concepts we explore target particular issues related to network representations and the general analysis of complex systems. However, our findings can also be of relevance to a broader scientific audience that is interested in various approaches to understand a given system. We classify the research presented in this thesis by *higher-order network models*. In this sense the term *higher-order* refers to the general inclusion of additional dimension of information that is available. In this way we expand the theory and methods of networks to be able to cope with the challenges induced by incorporating additional knowledge about a system.

In the following section we give a more detailed description of the exact problems we approach in this thesis.

1.2 Focus of the thesis

The focus of this thesis is the development of new models and the provision of novel perspectives in network theory that can be used to analyze complex systems. The framework of network theory is discussed in Chapter 2 and is used as basis for the investigations in the following chapters. In short, this theory captures interactions or relations, called *links*, between any kind of elements, entities or actors, called *nodes*. The discipline of networks science originated from mathematical graph theory and statistical physics but is also used in social sciences. Network theory is a quantitative approach to analyze any kind of complex systems. Further, properties can be imposed on the node and links to allow to represent a variety of complex systems as networks. Especially, we take into account additional dimensions of information that are often neglected in the commonly used network approach. In particular we investigate the aspect of temporality and interconnectivity which build the two main parts of the thesis.

Chapter 2 Before we focus on the specialized parts we give a brief introduction to the network framework and methodology. The definitions provided in this chapter build the basis for the later two main parts that extend and leverage on the fundamental network theory. We further introduce the framework of temporal and multi-layer networks as it is used throughout this thesis. There exist several definitions and notions that vary across different studies hence we clarify which approaches are used in our context. Additionally, both sections start with a brief overview of the motivation and advances in the corresponding fields of temporal and multi-layer networks.

In Part I: Temporality we present a model representation of temporal networks that allows to incorporate the order in which time-stamped links occur. We show that *ordering* of links can have high impacts on network measures and dynamic processes. In contrast to the *timing* of links the temporal order is often neglected in network studies. This is unjustified since certain order correlations can have strong implications on temporal networks. By order correlations we mean that the appearance of a link may depend on the ordering of appearance of previous links. What these implications are and in which situations the ordering should be considered is addressed and discussed in this part. We show that the temporal order of links can have a significant impact on importance measures and dynamical processes on the network. Comparing the findings of the model to actual temporal networks we highlight the advantage of our approach in capturing temporal characteristics.

Chapter 3 In the first chapter of Part I we introduce the framework of higher-order aggregate networks in terms of temporal networks. This modeling approach explicitly incorporates temporal order correlations expressed by path statistics in a network representation. We highlight the concept of *time-respecting paths* that is a fundamental component to capture the ordering of links. The so called *higher-order aggregate networks* explicitly capture path statistics of time-respecting paths and thus respect the ordering of links. This is a a general tool that can be applied to several kind of temporal systems. In particular, we investigate six empirical data sets, four of them consist of time-stamped links while the remaining two are based on path data.

Chapter 4 As a first application to address the impact of the temporal ordering of links we focus on centrality measures. We analyze three path-based centrality measures that are commonly used on *static* networks. We extend the definitions of these measures to develop *temporal* versions that can be applied to temporal networks. The temporal centrality measures consider the statistics of time-respecting paths rather than static aggregated paths. Additionally, we also define the centrality measures in the framework of higher-order aggregated networks. We analyze the various centralities on the six previously introduced data sets. We find that centrality measures based on our higher-order modeling framework captures better the true temporal centralities than measures based on a static time-independent approach.

Chapter 5 The temporal ordering of links effects *causality* of interactions and processes of temporal networks. By temporal causality we mean that a walk inside a network has to follow time-respecting paths in order to respect transitivity. Hence, in this chapter we analyze the influence of temporal order correlations on dynamical processes. In particular we focus on diffusion dynamics and therefore the spreading across a temporal network. Temporal causality is related to the ordering temporal links sequences, i.e. the existence of two links $a \to b$ and $b \to c$ does not necessarily imply the existence of a path $a \to b \to c$. Further, this is also related to the non-Markovian property of links sequences, meaning that the next link that is formed depends on the links that already appeared. So far it was usually assumed that considering the time-dependency of links slows down a dynamical process compared to a static analysis on a time-aggregated network. However, we show that both, a slow-down or speed-up, can result from the pure ordering of links. We quantify how strong temporal correlations are compared to a static representation and introduce a measure that allows to predict the change in diffusion speed. Further, we analytically discuss reasons for temporal order correlations and how reordering of links can slow-down or speed-up a process in a desired way.

In Part II: Interconnectivity we investigate model perspectives that comprise multiple networks that are *interconnected*. Considering multiple networks at the same time and considering links between them can have several advantages. It allows to study systems that are connected to each other or to encode different node and link types and the relation between them. There are various applications that can be comprised in the framework of *multi-layer networks*. To show the versatility of this perspective in this thesis we focus on two aspects that differ in their approach. On one hand we investigate multiple systems that are interconnected and the implications for a process running on the whole collection of systems. On the other hand we focus on a multi-dimensional system that consists of different types of relations and different types of elements that are connected to each other.

Chapter 6 In a first application we investigate systems where we only have limited knowledge about the link topology within or between different networks. In real world systems we often do not have complete knowledge of all of its components or subsystems. We may know that they are connected but may lack precise knowledge of the detailed link topology inside single parts. On the other hand we may know the precise link topology in single systems but may not know how precisely they are connected to each other. In this chapter we deal with the lack of knowledge in interconnected networks. We analytically assess how different topologies affect what can be said about a dynamical process in these situations. We use an ensemble approach to estimate the basic properties of a diffusion process and discuss in which cases the approximation is good enough and in which cases further information is needed.

Chapter 7 In a second application we study the impact of social mechanisms on citations of scholarly publications. The elements of this system are articles, authors and institutions which are connected by affiliation links. We analyze the citations between articles and how they translate to citations between scientists. In particular we compare how collaborative relations between scientists affect their citing behavior. Due to the vast amount of articles published in recent years we argue that scientist are more aware of their collaborators works which biases whom they cite. We quantify and measure the strength of this social effect on citations by analyzing different disciplines of physics journals. We further discuss how we can take into account the social influence to get new perspectives on scientific ranking schemes.

Even though, both parts of this thesis focus on two different aspects of complex systems they both share the feature of extending commonly used network approaches to include additional dimensions of information that are lost otherwise. In case of temporal networks

the additional dimension added is the time-dependence of links. The study of interconnected and multi-dimensional systems provides further contributions to the extension of the multi-layer framework. In the last chapter we summarize our findings and discuss future applications.

1.3 Contribution

Summarizing, this thesis contributes to network theory and its application to complex systems. As we outlined above understanding certain systems and deducing appropriate statements is not always a straightforward task. Hence, sometimes unnecessary restrictions and constraints are applied to analyze real-world systems with the known techniques and methods from network theory.

In this thesis we intend to overcome these limitations and discuss approaches to investigate complex systems in more appropriate manners. More precisely, we provide novel insights and methods to represent and analyze complex systems in a more comprehensive way. By focusing on the inclusion of temporal and multi-dimensional properties of networks and systems we contribute to the emerging fields of temporal networks and multi-layer networks. The methods and models are therefore a direct contribution to network theory and its applications.

Furthermore, our research targets real-world problems and provides solutions and insights to specific topics. In this way we also present results related to research questions from various fields in respect to particular real-world systems and applications that are not directly related to network science. Furthermore, the methods and models provided in this thesis can in general be applied to other systems that can be represented as networks. Therefore, our findings can be of interest to a broader variety of research topics that go beyond the mere network and complex systems analysis.

Chapter 2

Network Theory

Summary

In this chapter we introduce the basic definitions and methods of network theory. This includes notions and methodologies used throughout the thesis. We give a brief introduction of random walk processes that are applied to both temporal and multi-layer networks in later chapters. We further describe the application of statistical ensembles to networks that allow to deal with uncertainty and random graph models. Finally, in two separate sections we introduce and motivate the framework of temporal and multi-layer networks and how they are applied in the present work. We define temporal networks in a formal way and elaborate two alternatives how they can be represented for illustrative purposes. Multi-layer networks are described in a general way and the different types are briefly mentioned. The notions may differ to other works and applications and are therefore specified appropriately to our framework.

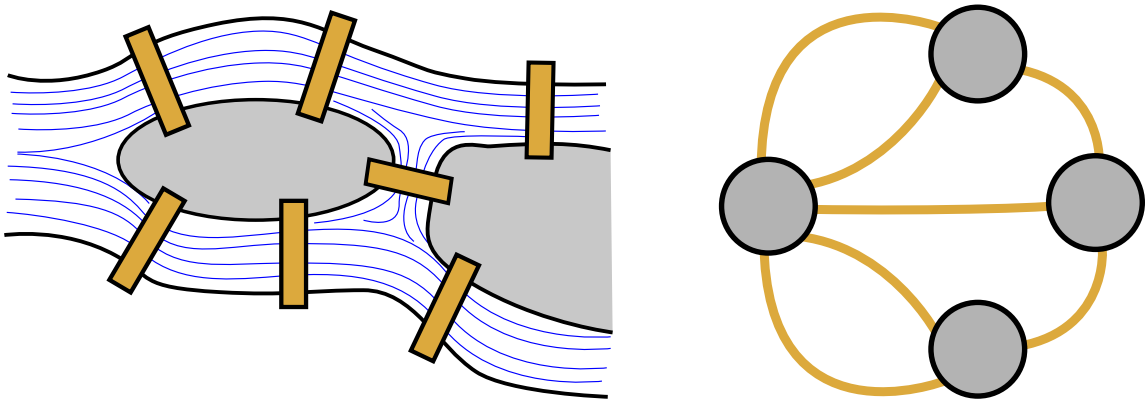
The framework presented in this chapter is based on common knowledge of network theory. The basic definitions can be found in a lot of works and are shaped for our purpose. The methods of temporal networks and multi-layer networks are also based on approaches used by other scientists. References are given for particular notions or definitions that are taken over. However, most of the theory is adjusted to our purpose and applications. Therefore some extensions are original for the work presented in the later chapters.

2.1 Introduction

Network science emerged from a lot of different disciplines [9, 111, 173, 176]. Networks are a useful representation of systems that consist of interrelated elements. As we have discussed in the introduction they are especially useful when dealing with complex systems. However, the theory itself was formally investigated even before it became popular in the applied sciences.

In mathematics network theory it is known as *graph theory* and is studied since a long time. The first contribution to this filed is assumed to be the famous Königsberg bridge problem that Leonhard Euler investigated in 1736. The question evolved around the seven bridges that connect Königsberg a city in former Prussia. The bridges crossed the Pregel River and allowed the citizens to access the different parts of the city. The city spread across two islands and the northern and southern part of the surrounding mainland splitting the city in four parts only connected by the aforementioned bridges. See Figure 2.1 for an illustration. Euler asked himself if it is possible to find a walk through the city that crosses each of the seven bridges exactly once. Further, it is imposed that there is no other way to cross the rivers and therefore only the streets of the city can be used to navigate. Since each of the four parts of the city allow to navigate freely the problem can be reduced to a simple *graph*. The four parts of the city can be represented by four *nodes* and the seven bridges by *links* that allow to traverse between the two nodes at its endpoints. The problems reduces to the task of starting at any node and traversing all seven links exactly once with the four node serving as access points to choose a bridge. Since one of the islands is connected with three bridges, a potential traveler would be forced to start or end his walk on this island. The reason is that the traveler has to use two bridges, one to leave and one to enter the island, leaving only one bridge to either enter or leave it a second time. However, the same also holds for the two mainlands of the city which both are only accessible by three bridges. This leads to an unsolvable problem since the traveler has to end or start his walk in three distinct parts of the city. Therefore Euler correctly concluded that there is no walk that crosses each of the seven bridges exactly once.

With his investigations Euler already explored some of the fundamental ingredients of graph respectively network theory, i.e. the relation of nodes and links and the formation of paths. The relationship of nodes and links connecting them can be projected to a lot of different scientific fields and problems. Nodes can represent any type of element or entity such as molecules, animals, countries, companies, websites and so on, the possibilities are endless. The same holds for the relations or links between these elements such as chemical reactions, mating behavior, trade agreements, financial liabilities or hyper-links.



(a) (b)

Figure 2.1: Königsberg bridge problem (a) Schematic representation of the seven bridges connecting the northern and southern mainland to the two island in the river. (b) Network representation where the nodes represent the mainlands, respectively islands, and the links represent the bridges.

Humans and their interactions with other humans such as friendships are a prominent filed of study in *social network analysis*. Some of the first examples of social networks were investigated by Jacob Moreno in 1934 [105]. Network theory not only allows to represent any kind of systems it also provides powerful tools to analyze them in a quantitative ways. Measures and techniques are continuously developed to cope with all the question and issues concerning networks.

However, not only *static* network that represent a particular instance of a system are of interest but also the time varying topology of interactions or processes running on them. to assume that links persist over time indefinitely is a naive assumption and does not align with most real world systems. The time-dependency of links can have a huge impact on several properties an metrics of networks and were and still are often neglected in network analysis. Further, network links usually only represent one type of interaction or connection implying the most networks only allow for one-dimensional link analysis. However, complex systems often exhibit different kind of interactions that can influences each other and the elements of the system. Such interaction dynamics are lost in standard networks that only capture a specific type of connection between its nodes.

In this chapter we will first introduce the basic methods and definitions of network theory, focused on the tools that are used in this thesis. In the following we will address some of the aforementioned limitations of classical network and provide extended frameworks to deal with time-dependent and multi-dimensional systems. The frameworks are essential for the approaches used in the later parts of the thesis that build on them.

2.2 Definitions

A *network* is a pair $G = (V, E)$ consisting of a set of *nodes* V and a set of *links* E. We call $V = V(G)$ the node set of G and $E = E(G)$ the link set of. The functions $V(\cdot)$ and $E(\cdot)$ are in general used to refer to the node and link set of a particular network.

The node set $V(G)$ is an unordered set of elements which comprise the nodes of the network. The number of elements in $V(G)$ denoted by $n = |V(G)|$ is called the *size*, *order* or *dimension* of G.

The link set $E(G)$ consist of pairs of nodes. More specifically,

$$E(G) \subseteq \{(a, b)|a, b \in V(G)\} = \binom{V(G)}{2}. \tag{2.1}$$

Therefore $E(G)$ is a subset of all possible combinations of two nodes of G. The nodes a and b of a link $e = (a, b)$ are called *endpoints*. The nodes that are part of a link e are *incident* to e. If a and b are incident to at least one common link they are *connected* and they are called *neighbors* or *adjacent* to each other. If nodes a and b are adjacent the relational symbol $\sim$ can be used, i.e. $a \sim b$. The number of links in $E(G)$ is often abbreviated by $m = |E(G)|$ A particular arrangements of links in a network is usually called network or link *topology*.

If the network is *undirected* the pairs in $E(G)$ are unordered and the links are called undirected. If the network is *directed* the pairs in $E(G)$ are ordered and the links are called directed. For a directed link $e = (a, b)$ we call a the *source* node and b the *target* node. To illustrate the link types a symbolic representation can be used. In case of a directed link we use $a \to b$ and for an undirected link $a - b$ or $a \leftrightarrow b$.

The network *density* refers to the ratio of the number of links present in the network to the maximal number of links that are possible. For undirected networks it is equal to $\frac{2m}{n(n-1)}$ and for directed network it is equal to $\frac{m}{n(n-1)}$.

The degree of a node v is the number of links incident to v and is denoted by $\deg(v)$. In case of a directed network we differentiate between *in-degree* and *out-degree*. The in-degree of an node is the sum of all links where v is a target node and the out-degree of node is the sum of all links where v is source node. The *total degree* of a node is usually the sum of the in- and out-degrees of the node.

A network is called *weighted* if links can appear multiple times and therefore $E(G)$ is a multi set. Otherwise the network is called *unweighted*. We denote the weight of a link $e = (a, b)$ by $\omega(e) = \omega((a, b))$. In case of weighted networks a *weighted degree* can be used

where the weight of links are also considered when calculating the degree.

A *path* is a ordered sequence of nodes $\{v_1, v_2, \ldots, v_k\}$ such that all consecutive nodes are part of a link in $E(G)$. More precisely, $(v_i, v_{i+1}) \in E(G)$ for all $i \in \{1, \ldots, k-1\}$. In case of a directed network the links have to be directed in the same way, i.e. v_i is the source node and v_{i+1} the target node for all the links that are part of the path. The *length of a path* is equal to $k-1$ which is the number of links that are part of the path.

A *shortest* path between two nodes a and b is a path of minimal length with a and b at its endpoints. Note that a shortest path does not have to be unique since there can be several paths with the same length. The legnth of a shortest path between two nodes a and b is also called *geodesic distance* or just distance between a and b. The *diameter* of a network is the maximal length of a shortest path between any two nodes, i.e. the maximal distance that occurs between two nodes.

A network is *connected* if any two nodes are part of at least one path. Otherwise it is called *disconnected*. In case of a directed network and if there exist a path between any two nodes that goes only in one direction that network is called *weakly* connected. If there exists a path in both directions for any two nodes then the network is called *strongly* connected.

A subnetwork G_s of G, denoted by $G_s \subseteq G$, is a pair $G_s = (V_s, E_s)$ such that $V_s \subseteq V$ and $E_s \subseteq E$. The connected subnetworks of a disconnected network are called *components* of G. The *largest connected component* of G is the component of maximal size and therefore contains the most nodes.

A useful way to represent a network is the *adjacency matrix*. It is equal to a matrix $\mathbf{A}$ of dimension n that is equal to the number of nodes in the network. As indicated by the name the adjacency matrix captures which nodes are adjacent to each other. For a unweighted network the elements $A_{i,j}$ of $\mathbf{A}$ are equal to 1 if there exist a link between the nodes i and j, otherwise $A_{i,j}$ is equal to zero. Note, that the adjacency matrix requires that the nodes are labeled by numbers or ordered in such a manner that they can be attributed to rows and columns of the matrix. In case of an undirected network $\mathbf{A}$ is symmetric and therefore $A_{i,j} = A_{j,i}$. A weighted adjacency matrix can capture the weights of links such that $A_{i,j} = \omega((i,j))$ with $\omega((i,j)) = 0$ if there is no link between i and j.

In mathematics networks are called *graphs*, the nodes are called *vertices* and the links are called *edges*. The other notions discussed in this sections generally apply to both, network and graph theory.

2.2.1 Random walks

Here we introduce the framework of a *random walk process* that runs on the topology of a network. In the following we provide a short summary that highlights the important properties of this process. Note that these are facts already known from the theory of random walks [16, 93], but later on we will aplly this framework to temporal and multilayer networks.

We assume a discrete time random walk process on a network $\mathbf{G}$ that consist of n nodes. Starting at an arbitrary node, at each step of the process the walker moves to an adjacent node. For a pair of nodes i, j the probability $P(i \to j)$ for a walker to move from node i to node j is given by the corresponding entry $\mathbf{T}_{ij}$ of a *transition matrix* $\mathbf{T}$. Since we have $\sum_j P(i \to j) = 1$, the transition matrix is row stochastic.

We further consider a vector $\pi^t \in \mathbf{R}^n$, whose entries π_i^t indicate the probability of a random walker to visit node i after t steps of the process. Here, we consider π^0 as a given initial distribution, whose entries π_i^0 give the probability that the random walker has started at node i. The change of visitation probabilities $\pi^t \to \pi^{t+1}$ can then be calculated based on the transition matrix as follows:

$$\pi^{t+1} = \pi^t \mathbf{T}. \tag{2.2}$$

Since this is an iterative process starting with π_0, the visitation probability vector after t time steps can be calculated as $\pi^t = \pi^0 \mathbf{T}^t$, and we can investigate the long-term behavior of the random walk process for $t \to \infty$. For a visitation probability vector π^* such that $\pi^* \mathbf{T} = v^*$, we can say that the process reaches a stationary distribution π^*, and if the transition matrix $\mathbf{T}$ is primitive, the Perron-Frobenius theorem guarantees that such a unique stationary distribution π^* exist.

In order to assess the convergence time of a random walk process, we can study the *total variation distance* between visitation probabilities π^t after t steps and the stationary distribution π^*. For two distributions π and π', the total variation distance is defined according to Ref. [138] as

$$\Delta(\pi, \pi') := \frac{1}{2} \sum_i |\pi_i - \pi_i'| , \tag{2.3}$$

where π_i indicates the i-th entry of π.

As a proxy for diffusion speed, we can now investigate how long it takes until the total variation distance $\Delta(\pi^t, \pi^*)$ falls below some given threshold value ε. In other words, we study how many steps $t(\varepsilon)$ a random walker needs such that $\Delta(\pi^t, \pi^*) \leq \varepsilon$ for $t \geq t(\varepsilon)$.

The eigenvalues $1 = \lambda_1 \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$ of a row-stochastic matrix necessarily have

absolute values that fall between zero and one, while the largest eigenvalue λ_1 is necessarily one. The number of required time steps $t(\varepsilon)$ (and thus the diffusion speed of the random walk process) can be estimated by means of the the second-largest eigenvalue λ_2 of $\mathbf{T}$,

$$t(\varepsilon) \sim \frac{-1}{\ln(|\lambda_2|)}\,. \tag{2.4}$$

For a detailed derivation see Ref. [27]. Eq. (2.4) shows that a second-largest eigenvalue λ_2 close to one implies slow convergence, while λ_2 close to zero implies fast convergence. Therefore in the following we use the second-largest eigenvalue of a transition matrix $\lambda_2(\mathbf{T})$ as a proxy to measure and quantify the convergence behavior of a random walk on a network.

2.2.2 Statistical ensemble

In network theory we can roughly distinguish two kind of perspectives, a *micro* and a *macro* perspective. The micro perspective includes everything that is focused on a *particular network*. Such as the degree of nodes or more general network measures such as the diameter, the modularity and other topological indicators.

On the other hand the macro perspective only deals with statistical characterizations. It provides no exact information about single nodes or particular networks. One rather relies on stylized facts that comprise empirical findings. Hence, this perspective deals with *network classes*.

To analyze classes of networks a stochastic approach is needed that deals with the lack of knowledge about the microscopic details of a network. First some known properties or aggregated statistics are considered as fixed. For example, this can be the number of nodes and links or the degree distribution. Then all possible realizations of networks are studied that share the same fixed properties. The set of all realizations preserving some statistical properties is called an *ensemble*.

Formally, we denote the fixed aggregate statistics or properties of a network ensemble by the *macro state* X. Based on stochastic model $\mathcal{M}(X)$ we can generate network realizations that are consistent with the macro state X. All network realizations that are consistent with X are part of a sample space $\Omega(X)$. This means each network $G \in \Omega(X)$ represents a *micro state* that is consistent with the macro state X. We can define a probability measure P on the sample space Ω to assign a probability measure to each micro state $G \in \Omega(X)$. We denote the ensemble of all realization given a stochastic model $\mathcal{M}$ that are consistent with X by $\mathcal{E}(\mathcal{M}; X)$ or $\mathcal{E}(\mathcal{M}(X))$.

The goal of this construction is to study expected properties of network ensembles.

A reverse approach is to find an appropriate stochastic model that explains a particular network in the best way. Assume that we observe an empirical network G_e. Given a stochastic model $\mathcal{M}$ we can compute the *likelihood* that G_e was generated by $\mathcal{M}$,

$$\mathcal{L}(G_e|\mathcal{M}) = \mathcal{L}(\mathcal{M}) := P(G_e|\mathcal{M})\,. \tag{2.5}$$

Hence the likelihood is the conditional probability of G_e given the stochastic model $\mathcal{M}$. A model is most plausible to generate G_e if its maximizes its likelihood. Therefore we intend to find the maximum likelihood estimator $\hat{\mathcal{M}}$,

$$\hat{\mathcal{M}} = \text{argmax}_{\mathcal{M}}\mathcal{L}(\mathcal{M})\,. \tag{2.6}$$

This procedure is equivalent to finding the most likely statistical ensemble for G_e and is also called *statistical inference*. The ensemble that maximizes the conditional probability encodes knowledge about the empirical network G_e.

Erdös-Rényi network A direct application of statistical ensembles to network theory are *random graph models*. The most popular one is the *Erdös-Rényi model (ER)* [37] usually denoted by $G(n, m)$ or $G(n, p)$. The parameters n and m respectively p represent the macro state X of the stochastic model. The parameter n denotes the numbers of nodes that is given. There are two version of the Erdös-Rényi model, one additionally considers the number of links m the other one the probability p that there is a link between any two nodes. The $G(n, m)$ model uniformly at random chooses m pair of nodes and generates a link. The $G(n, p)$ model generates a link with probability p for each pair of nodes. For n converging to infinity the two version are approximately the same. The ensembles can be denoted as $\mathcal{E}(G(n, p))$ and $\mathcal{E}(G(n, m))$. All network realization that are part of this ensemble share the same number of nodes n and the same number of links. Hence the Erdös-Rényi model can be used to randomly generate network realizations that all share the same link density.

2.3 Temporality

The network perspective provides a useful framework to study structures and dynamics of complex systems. It is based on connectivity, interactions or other kind of relations between elements in a system. However, sometimes not all dimensions of knowledge or information of the system can be represented by the basic framework. In this situations and extended modeling perspective is need to incorporate the desired properties or relation. One of the most fundamental additional dimensions that is present in almost all natural systems and real world data is *time*. Interactions or connections in a system usually do not happen all at once and are also not persistent for eternity. To deal with the consequences of time as an additional dimension an extended perspective is needed. One approach to address this are *temporal networks*, a network perspective that includes time as an additional ingredient in various ways.

The time dimension is present in most real-world data and there are different ways to deal with it. Usually network studies got rid of the time dependency by aggregating interaction happening in the system over some period of time. By summing up all interactions that happen during a particular *time-window* the data get aggregated to allow for a static representation. Link weights can be used to indicate how often a link $a - b$ was present over the time horizon of interest. If the data is spread over a large period of time, multiple time-windows can be defined that slice the system into several instances. The single *time-slices* can then be analyzed separately allowing to study the time evolution of certain network measures. However, not all aspect of temporal structures, dependencies and dynamics can be captured in this way.

Consider for example a transportation system that is prone to correct scheduling and rely on fast delivery of needed goods. Assume that we need to send a shipment from harbor A to harbor C, given that there is no cargo route from A to C and therefore no direct way to pass on the delivery. However, assume there is an intermediate harbor B that maintains shipping routes to harbor A and C. From a static network perspective the problem would be solved by sending the shipment from A via B to C. However, in reality there are some crucial issues that have to be considered when planning the execution of the delivery. It is often important that the *fastest* way of delivery is chosen since delaying the transportation can have severe monetary consequences. There may be different routes connecting A to B and B to C that take different amount of time depending on the vessels that are used or the amount of stops in between. Further, we have to assure that the delivery arrives in harbor B before the connecting ship to C leaves it. This implies some *waiting time* of the shipment in harbor B before the departure. Therefore, an optimized planing of a

transport requires more detailed information about the system then what is captured in a static network [15, 178].

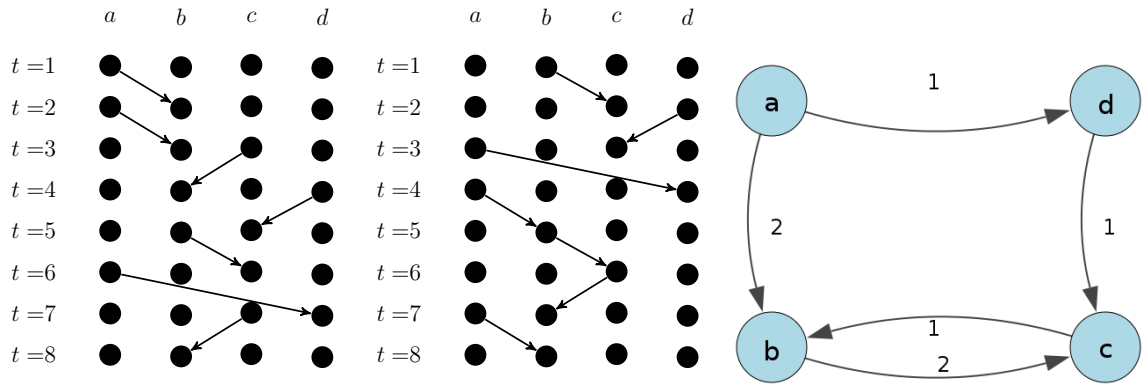
The example shows that the *timing* and *ordering* of links can be crucial in a temporal networks. Transportation systems, such as air transportations [119, 140], are not the only application the requires the inclusion of time as a crucial dimension. Communication and contacts between humans are a prominent example that includes several temporal dimensions that should be considered when studying human interactions [34, 67, 89, 181]. Especially, in spreading of information and disease the temporal component can play an crucial part [3, 70, 165]. In this thesis will we in particular focus on four interaction systems and two transportation systems. But there are plenty other applications in biological systems, such as cell biology and neural networks but also in ecological networks and population biology. We refer the readers to two review articles by Petter Holme that capture the diversity of fields dealing with temporal networks [68, 69].

The framework of temporal networks differs for the various works on the topic and are usually tailored in respect of the property under consideration. In the following section we introduce the definitions and methods that we use through out the thesis and represents our perspective on temporal networks.

2.3.1 Temporal networks

In the classical sense networks consist of nodes and links. If needed one can also assign properties to the nodes and links such as link weights. However, to capture the temporality of links a richer framework is needed. We start by formally defining what a temporal network is and how we describe it. We present how they can be visualized and what means of aggregation can be used to simplify them.

In the following we clarify what we mean by the notion of a *temporal network*. We define a temporal network $G^T = (V, E^T)$ as a tuple consisting of a set of nodes V and a set $E^T \subseteq V \times V \times [0, T]$ of time-stamped links $(v, w; t) \in E^T$ for an observation period $[0, T]$. The nodes have no temporal component and persist over time, only the links are considered to be time-dependent. It is important to note that we assume *discrete* time stamps $t \in [0, T]$. This implies that we can not directly assign a *duration* to a link (v, w). A time stamped link $(v, w; t)$ indicates only the *instantaneous* presence of the link (v, w) at time t. If a link (v, w) is present for some time interval $[t_{start}, t_{end}]$ we can use a small unit of discrete time Δ and add multiple time-stamped links $(v, w; t)$ at time stamps $t = t_{start}, t_{start} + \Delta t, t_{start} + 2\Delta t, \ldots, t_{end}$. Even though, the discreteness assumption does not allow for continuous links activity it can naturally be applied to real-world data.



(a) Temporal network G_1 (b) Temporal network G_2 (c) Weighted, time-aggregated representation of both G_1 and G_2

Figure 2.2: Illustration of temporal networks Time-unfolded and weighted static, time-aggregated representation of two temporal networks G_1 and G_2.

Time-stamped data is typically obtained based on some sort of *sampling*, whose sampling frequency defines the smallest unit of time Δt.

2.3.2 Time-unfolded and time-aggregated networks

It can be helpful to visualize a temporal network G^T. To disentangle the temporal structure that is contained in the link set E^T we use the so-called *time-unfolded networks*. It is a simple two-dimensional static representation. All nodes are arranged on a horizontal dimension while time is unfolded to an additional vertical dimension as illustrated in Figure 2.2. For an observation period $[0, \ldots, T]$ and a given Δt we can then add *temporal copies* of all nodes for all possible time steps $k\Delta t$ (for $k = 0, 1, \ldots$). For simplicity, in the following we assume $\Delta t = 1$, which allows us to denote the temporal copies of a node v as $v_t, v_{t+1}, v_{t+2}, \ldots$. The main benefit of this construction is that it allows us to represent a time-stamped link $(v, w; t)$ by means of a static link (v_t, w_{t+1}) connecting the temporal copies v_t and w_{t+1} of node v and node w respectively. The intuition behind this notation is that a quantity residing at node v at time t can move to node w via a time-stamped link $(v, w; t)$, arriving there at the next time step $t + 1$. Two simple examples for time-unfolded static representations of two different temporal networks with four nodes and seven time-stamped links are shown in Figure 2.2a and Figure 2.2b.

Despite the recent development of methods to study temporal networks, the most widespread way to study time-stamped network data is to aggregate all time-stamped links into a static, *time-aggregated network* $G = (V, E)$. This means that, given a temporal

network $G^T = (V, E^T)$, two nodes $v, w \in V$ are connected in the static network whenever a time-stamped link exists at *any* time stamp, i.e., $(v, w) \in E$ if and only if $(v, w; t) \in E^T$ for any $t \in [0, T]$. Additional information about the statistics of time-stamped links in the underlying temporal network can be preserved by considering a *weighted time-aggregated* network, in which weights $\omega(v, w)$ indicate the number of times time-stamped links $(v, w; t)$ have been active during the observation period. I.e., we consider a weighted time-aggregated network with a weight function $\omega : E \to \mathbb{N}$ defined as

$$\omega(v, w) := |\{t \in [0, T] \,|(v, w; t) \in E^T\}|.$$

Figure 2.2c shows the weighted, time-aggregated networks corresponding to the two temporal networks shown in Figure 2.2a and Figure 2.2b. These simple examples highlight the important observation that *different* temporal networks are consistent with the *same* weighted, time-aggregated network. This is due to the fact that in the time-aggregated network we lose all information on both the timing and the ordering of links in the temporal network.

In Part I of the thesis we will further investigate the relevance *ordering* in temporal networks. In Chapter 3 we discuss in more detail how one can navigate inside a temporal network and how temporal paths can be defined. Nevertheless, the framework presented so far already allows to capture time-stamped links in a network representation. This extension only affects the link set E^T that considers an additional parameter t for each link indicating the point in time where the link was present. However, it is important to emphasize that temporal network framework that we consider in this thesis only considers *discrete* time-stamps. Other approaches on temporal networks may use different notations or frameworks in this respect.

2.4 Interconnectivity

Networks provide as useful tool do represent and analyze a variety of systems. Sometimes they have to be adjusted or extended to capture all of the relevant system properties that one is interested in. One particular extension is the inclusion of time which lead to temporal networks, as discussed in the previous section. Further, systems are often not independent of their surrounding and can be influenced and therefore be connected to other systems. This *interconnectivity* between systems also adds an additional dimension to the system. The general inclusion of multiple dimensions and different networks in one framework is referred to as *multi-layer networks*.

One of the first examples studied in terms of interconnected or *layered* systems was a power grid that revealed the importance of a comprehensive representation of complex system to prevent cascades of failures [137]. The investigation evolved around an electrical blackout that occurred in September 2003 in Italy. The blackout was sever and affected all of the Italian mainland for several hours. The trigger was a cut of a power line, caused by a storm, that supplied Italy from Switzerland. This disturbance lead to the failure of other power supply lines from France which were exposed to an increased demand. The circumstances that lead to the blackout had a technical origin and initially only affected the northern part of Italy. However, also the communication system SCADA that is used to manage the electric power grid was affected by the power shortage. As a consequence the organization maintaining the power grid lost control and could not intervene. This lead to the failure of further power stations increasing the area of the blackout. The SCADA system itself is based on several communication servers throughout Italy, so that the failure of one communication node should not impact the whole systems. However, the failure of additional power stations also lead to the failure of further communication servers that relied on them. The dependency between power station and communication servers used to operate them were constructed in such a way that the blackout cascade took all over Italy. The failure of a power station lead to the failure of a communication server that in reverse affect further power stations and so on. This *inter-dependency* of the power grid and the communication network was not foreseen. The two systems were regarded in a independent way even though they crucially were dependent on each other.

The power grid example tells us that an isolated view on a complex systems can lead to undesired consequences simply by intentionally or unintentionally neglecting the interconnectivity between two systems. Cascading effects and interdependency networks are nowadays a well studied subject [23, 46, 179]. Interacting and interconnected systems are therefore the focus of several studies [4, 87, 91, 149, 166]. The generalization of interconnected network led to the investigation of *network of networks (NON)* [29, 31, 45, 94, 122]. In this perspective the interaction and dynamics inside a single network are secondary rather the iteration between networks are focused on.

Multi-dimensional relations on the same system can also be studied from the perspective of interconnected networks. Rather then focusing only on one type of interaction or connection between elements of a systems several type of ties are investigated. Such kind of studies are especially prominent in social network analysis were one tie between a pair of individuals is usually not enough to understand and analyze social behavior and relationships. One of the first investigations in this direction were done by Jacob Moreno in 1934 [105]. He studied the relationship between members of a cottage family which he depicted as a network. However, not only the existence of a social tie between family

members were indicated but also different *types* such as friendly and refusing ties. This way he already captured several dimensions in one network representation. Different type of links are not only useful in social systems but also in other applications [25, 44, 76, 90].

Another approach to study multi-dimensional relations in a network is based on the theory of *hypergraphs* [14]. They can be used in various applications [43, 53] and consider that a single links does not only connect two nodes but a set of nodes. Depending on the real-word system under investigation this methodology can offer a complementary perspective.

Until recently, the standard approach in the literature considered networks as isolated entities that do not interact with other networks. Today we understand that this assumption is a rough simplification, since real networks usually have complex patterns of interaction with other networks. In order to study more realistic systems, network theory extended its perspective to account for these network to network interactions, and to investigate their influence on various processes of interest that may use the network topology as substrate [46, 47, 55, 123]. Networks consisting of multiple networks and the connections between them are called *interconnected* or *multi-layer* networks.

2.4.1 Multi-layer networks

In the following section we introduce the definitions and notions of multi-layer networks and how they are applied in this thesis. Similar frameworks are commonly used nowadays, see for example Ref. [18, 82].

We define a multi-layer network as a tuple $\mathbf{M} = (\mathbf{G}, E_I)$. The *layer* set $\mathbf{G}$ is a set of L *layers* $G_1, \ldots, G_L$. Each of these layers G_l is a single-layer network $G_l = (V_l, E_l)$ where $V(G_l)$ and $E(G_l)$ denote the nodes and links of layer l respectively. We call the links $E(G_l)$ between nodes *within* the layers l *intra*-links. E_I is the set of *interconnectivity* links between nodes that are part of different layers. We can formalize it as,

$$E_I = \{E_{s,t} \subseteq G_s \times G_t | s, t \in \{1, \ldots, L\}, s \neq t\}\,, \tag{2.7}$$

where $E_{s,t}$ contains all links that connect nodes from G_s to nodes in G_t. We call all links that are part of E_I *inter*-layer links, i.e. all links (u, v) for which $u \in V(G_s)$ and $v \in V(G_t)$ for $s \neq t$. Inter-layer links induce a multipartite network with the independent sets $G_1, \ldots, G_L$.

The previous definitions are only a basic notion for a general framework. Depending on the application the meaning of intra-layer and inter-layer links can vary. For example in Chapter 6 inter-links and intra-links are only distinguishable in regard to the nodes

they connect. Links that connect nodes from the same layer are considered intra-links and links that connect nodes from different layers are considered inter-links. In respect to their *functionality* in regard of a process or meaning they do not differ. However, in Chapter 7 inter-layer links are correspondence links that connect nodes in one layer to different type of nodes *affiliated* to them in another layer. On the other hand intra-layer links are relational links between nodes of the same type.

Already this two examples highlight the huge variety of applications of the multi-layer network approach. Therfore in each case the meaning of layers, inter- and intra-layer links and different kind of nodes have to be clarified.

2.4.2 Types of multi-layer networks

As mentioned before the formal definitions presented in the previous section allow for different interpretations. Depending on the application there are various possibilities to represent multi-dimensional data in a multi-layer framework. However, some of them share common properties that allow to classify or categorize them. In the following we briefly present commonly used notions that specify particular representations of multi-layer networks.

Network of networks To study interactions between different systems and how their properties affect each other it is sometimes useful to regard them as *networks of networks*. This notion implies that we take the perspective that each single network is regarded as node in a meta network that connects the single networks. In such cases the network layers usually share common properties or are of the same type such that the links within layer have the same meaning. Hence, the focus in networks of networks are the inter-layer links. The intra-layer topology is usually reduced to aggregate measures that allow to represent this layer as a single node in a network of networks. Further, also the inter-layer links are aggregated in such a way that only the layer that get connected are of interest and not single nodes. Therefore the network of networks perspective is taken to investigate global mechanisms that act on the whole multi-layer system and where the properties of single layers are only relevant in regard to their contribution to the whole system. A network of networks usually loses some of the information that is available in single layers and therefore is a aggregation of a multi-layer system to a single network. An example of such a *multi-layer aggregation* is presented in Chapter 6.

Multiplex network A common way to represent different kind of relations between the same set of nodes are so called *multiplex* networks. They build a subclass of general multi-layer networks where the set of nodes is the same for all layers. Formally, $V_l = V$ for all $l \in \{1, \ldots, L\}$. Therefore, a particular node v exists L times in the multiplex networks, once as copy for each layer $v_1, \ldots v_L$. The only thing that differs between the layers are the intra-layer links. The inter-layer links are only considered to be correspondence links that connect the same node throughout all layers. The set of inter-layer links of a multiplex network can be defined as follows,

$$E_I = \{(v_i, v_j) | v \in V, i, j \in \{1, \ldots, L\}, i \neq j\} \,. \tag{2.8}$$

If the layers are stacked in a specific order it is sometimes implied that only nodes of consecutive layers are connected such that

$$E_I = \{(v_i, v_{i+1}) | v \in V, i, \in \{1, \ldots, L-1\}\} \,. \tag{2.9}$$

Multiplex networks are well suited for application where different types of links between the same set of nodes are investigated. An application of a multiplex network representation is used in Chapter 7.

Multiplex networks can also be aggregated to a single layer which is then called a *monoplex network*. The monoplex $G = (V, E)$ consist of the initial node set and the aggregation of all the intra-layer links

$$E = \bigcup_{l=1}^{L} E_l \,. \tag{2.10}$$

The monoplex perspective is therefore the aggregated view that does not allow to distinguish layers. It can be used to represent all of the relations between nodes with out taking into account different types of links.

Part I

Temporality

"The whole history of science has been the gradual realization that events do not happen in an arbitrary manner, but that they reflect a certain underlying order, which may or may not be divinely inspired."

Stephen Hawking
A Brief History of Time (1988)

Chapter 3

Temporal Ordering

Summary

Despite recent advances in the study of temporal networks, the analysis of time-stamped network data is still a fundamental challenge. In particular, recent studies have shown that correlations in the *ordering of links* crucially alter *causal topologies* of temporal networks, thus invalidating analyses based on static, time-aggregated representations of time-stamped data. These findings not only highlight an important dimension of complexity in temporal networks, but also call for new network-analytical methods suitable to analyze complex systems with time-varying topologies. Here we introduce a modeling framework that allows to capture the ordering of links in temporal networks. Our approach demonstrates that higher-order aggregate networks constitute a powerful abstraction, with broad perspectives for the design of new, computationally efficient data mining techniques for time-stamped relational data.

Based on the framework of higher-order aggregate networks introduced in Scholtes, I., Wider, N., Pfitzner, R., Garas, A., Tessone, C.J. and Schweitzer, F. *Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks*, Nature Communications, vol. 5, number 5024, 2014. and Scholtes, I., Wider, N., Garas, A. and Schweitzer, F. *Higher-Order Aggregate Networks in the Analysis of Temporal Networks: Path structures and centralities* , Eur. Phys. J. B 89 (3) 61, 2016. NW contributed to the development, refinement and application of the methodology. Further, NW analyzed the data in perspective of the higher-order approach.

3.1 Introduction

The network perspective has provided valuable insights into the structure and dynamics of numerous complex systems in nature, society and technology. However, most of the complex systems studied from this perspective are not static, but rather exhibit time-varying interaction topologies in which elements are only linked to each other at specific times or during particular time intervals. The increasing availability of high-resolution data on time-stamped or time-ordered interactions from a variety of complex systems has fostered research on how different aspects of the temporal dynamics of networks influence their properties. While the *topological* characteristics resulting from which elements are linked to which other elements have been studied extensively, the importance of the additional *temporal* dimension resulting from *when* these links occur has become clear only recently. Despite an increasing volume of research, its full impact on the properties of complex systems and on the evolution of dynamical processes still eludes our understanding [68, 69]. Addressing this open issue, different strands of research have focused on the question how different types of temporal characteristics of complex networks – such as the activation times of nodes, the inter-event times between links, the duration and/or concurrency of interactions, or the order in which these interactions occur – affect the properties of temporal networks as well as dynamical processes evolving on them. Assuming that network topologies change in response to the dynamical process running on top of them, another line of research has studied adaptive networks, again highlighting that network dynamics have important consequences for dynamical processes [59, 60].

Apart from the *timing* of interactions, the *order* in which these interactions occur is another important characteristic of temporal networks. Compared to the rich literature on node activities, a relatively smaller number of studies empirically investigated effects of causality in temporal networks [80, 84, 85, 86, 88, 127, 135, 139]. Not only does the ordering of interactions crucially affect causality in temporal networks, it has also been shown to dramatically shift the evolution of dynamical processes compared to what we would expect based on a static, time-aggregated perspective [86, 88, 127, 140, 144, 158]. Some of these works have further taken a modeling perspective, highlighting that real-world temporal network data exhibit non-Markovian characteristics in the sequence of links which are not in line with the Markovianity assumption that is (implicitly) made when studying static representations of time-varying complex networks. Neglecting these non-Markovian characteristics not only leads to wrong results about dynamical processes, it also leads to wrong centrality-based rankings of nodes, as well as misleading results about community structures [86, 115, 140, 144].

The main reason why an analysis of static, time-aggregated networks yields misleading results about the properties of temporal networks is that the ordering of links can alter path structures in temporal networks compared to what we would expect based on their static topology. Precisely, in static networks the presence of two links (a, b) and (b, c) connecting nodes a to b and b to c necessarily implies that a path from a via b to c exists. However in a temporal network, for a to be able to influence c the link (a, b) must occur *before* the link (b, c) and thus the presence of a path depends on the ordering of links. The so-called *time-respecting paths* respect causality, i.e. a time-respecting path only exists if link (a, b) occurs *before* link (b, c) [69, 80]. In order to additionally consider the *timing* of interactions, it is common practice to impose the additional constraint that links (a, b) and (b, c) must occur within a certain time window, thus imposing a limit on the time a particular process can wait in node b. As such, both the *order and timing* of interactions affect time-respecting paths - and thus causality - in temporal networks. This simple example highlights that the mere ordering of links in temporal networks can introduce an additional temporal-topological dimension that can neither be understood from the analysis of static, time-aggregated representations, nor from the analysis of inter-event times or node activity distributions [127].

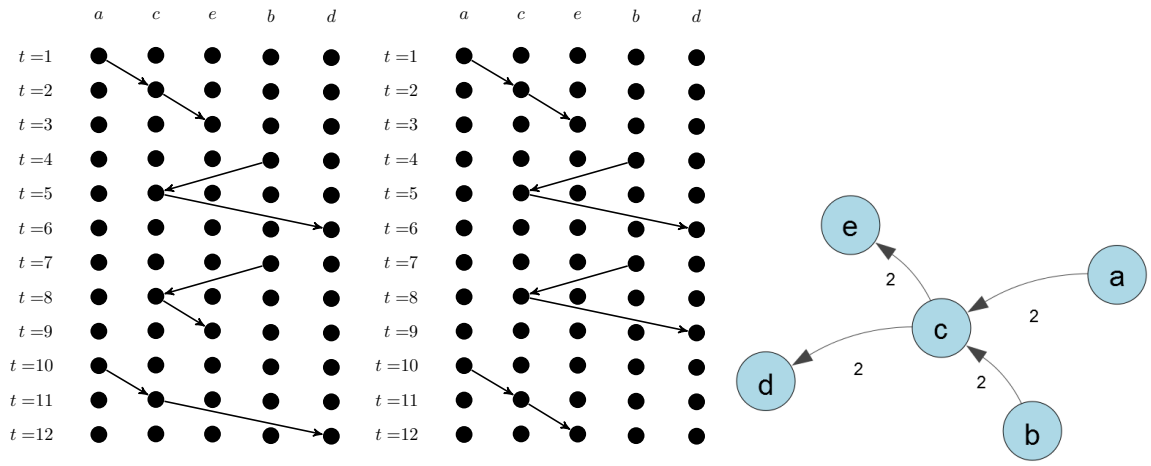
The remainder of this chapter is structured as follows: In section 3.2 we first introduce basic concepts of time-respecting paths with maximum time differences between consecutive links. In section 3.3 we introduce the framework of higher-order time-aggregated networks, a simple abstraction of temporal networks that takes into account the statistics of time-respecting paths up to a given length. In section 3.4 we present six empirical data that exhibit temporal path structures and are analyzed by our framework in the following chapters. This chapter introduces the fundamental concept and framework that will be applied to several issues and real-world systems in Chapter 4 and Chapter 5

3.2 Time-respecting paths

The *ordering* of time-stamped links plays a crucial role in respect of causality of temporal interactions. In this section we further investigate the concept of paths that respect the ordering of links.

Recall the definition of a temporal network given in section 2.3.1. A temporal network $G^T = (V, E^T)$ is a tuple consisting of a set of nodes and a set of time-stamped links for a observation period $[0, T]$. Also recall that in our framework a time-stamped link $(v, w : t)$ exists instantaneously at a discrete time step t.

Importantly, both the timing and the ordering of links influence path structures in tem-



(a) Temporal network G_1 (b) Temporal network G_2 (c) Weighted, time-aggregated representation of both G_1 and G_2

Figure 3.1: Illustrative examples Time-unfolded and weighted static, time-aggregated representation of two temporal networks G_1 and G_2

poral networks. In particular, in the context of temporal networks we must consider *time-respecting paths*, an extension of the concept of paths in static network topologies which additionally respects the timing and ordering of time-stamped links [69, 79, 119]. In this thesis we define a time-respecting path between a source node v and a target node w to be any sequence of time-stamped links

$$(v_0, v_1; t_1), (v_1, v_2; t_2) \ldots, (v_{l-1}, v_l; t_l)$$

such that $v_0 = v, v_l = w$ and the sequence of time-stamps is increasing, i.e. $t_1 < t_2 \ldots < t_l$. The latter condition on the ordering of links is particularly important since it is a necessary condition for causality. This means that for any temporal network a node a is able to influence node c based on two time-stamped links (a, b) and (b, c) only if link (a, b) has occurred *before* link (b, c). A simple example for a time-respecting path $(a, c; 1)$, $(c, d; 5)$ can be seen in Figure 3.1a, where the time-unfolded representation of the temporal network G_1 is illustrated.

At this point, it is important to note that, different from the usual notion of paths in static networks, the question whether a time-respecting path exists between two nodes requires to specify a *start time* $t_0 \leq t_1$. In the example of Figure 3.1a we observe a time-respecting path $(a, c; t_1 = 1), (c, d; t_2 = 5)$ between node a and d, which can only be taken if we consider paths starting at node a at time $t_0 = 1$. If instead we were to ask for a time-respecting path between a and d starting at node a at time $t_0 = 5$, our only choice

would be the path $(a, c; 10), (c, d; 11)$.

3.2.1 Maximum time difference

In the definition of a time-respecting path above, we have required that the sequence of time stamps of the links constituting the path must be increasing. Clearly, this condition is rather weak since it makes no assumptions whatsoever about the time difference between two consecutive time-stamped links on a time-respecting path. As such, for the mere existence of a time-respecting path in a temporal network evolving over a period of years, it is actually not important whether the time difference between two consecutive links is a few seconds or a few years.

However, we typically study time-respecting path structures because they constitute the substrate for the evolution of dynamical processes which have intrinsic time scales that are much smaller than the period during which we observe a temporal network. In the study of time-respecting paths, it is thus often reasonable to impose a *maximum time difference* δ, i.e. we limit the temporal gap between two consecutive time-stamped links that are considered to contribute to a time-respecting path to a maximum of δ [69, 119]. In this case, rather than requiring a mere increasing sequence of time-stamps, we demand that the condition $0 < t_{i+1} - t_i \leq \delta$ must be fulfilled for all $i = 1, \ldots, l-1$. For a maximum time difference of $\delta = 1$, we thus limit ourselves to the study of time-respecting paths for which all time-stamped links occur at immediately consecutive time stamps. As another limiting case, we can consider $\delta = \infty$, which means that we impose no further condition apart from the requirement that the sequence of time stamps of links on a time-respecting path is increasing. Revisiting the example of Figure 3.1a, we observe that the time-respecting path $(a, c; 1), (c, d; 5)$ only exists if we allow for a maximum time difference $\delta = 4$, while for all $\delta < 4$ the only time-respecting path between the nodes a and d is $(a, c; 10), (c, d; 11)$.

3.2.2 Shortest and fastest time-respecting paths

Let us now formally define the length of time-respecting paths in a temporal network, which will allow us to define the notion of *shortest time-respecting* paths used throughout our work. Due to the additional temporal dimension, the length of a time-respecting path

$$(v_0, v_1; t_1), \ldots, (v_{l-1}, v_l; t_l)$$

can be studied both from a topological and a temporal perspective. Following the usual terminology, we call the number l of time-stamped links on a time-respecting path the

(topological) *length* of the path. We further call the time difference $t_l - t_1 + 1$ the *duration* of the path. Here the increment by one accounts for the duration of the final link $(v_{l-1}, v_l; t_l)$, i.e. for the fact that any process starting at node v_0 at time t_1 will only reach node v_l at time t_{l+1}.

Having defined both, the length and duration of time-respecting paths, it is now trivial to define the *shortest time-respecting path* between two nodes v and w as the time-respecting path with the smallest (topological) length. In analogy, we define the *fastest time-respecting path* as the time-respecting path with the smallest (temporal) duration. Following our previous comment about the necessity to define a start time t_0 for a time-respecting path, it is clear that the shortest or fastest time-respecting path can only be found unambiguously with respect to a given start time t_0, i.e. at different times during the evolution of a temporal network the same pair of nodes can be connected by different shortest or fastest time-respecting paths.

3.2.3 Transitivity of paths in static and temporal networks

Let us conclude this preliminary section by highlighting important differences between paths in static networks compared to time-respecting paths in temporal networks, that result from the ordering and timing of links. Let us first highlight that paths in static networks are *transitive*. This means that from the presence of two paths $(v_0, v_1), \ldots, (v_{k-1}, v_k)$ and $(v_k, v_{k+1}), \ldots, (v_{l-1}, v_l)$ between v_0 and v_k and between v_k and v_l respectively, we can conclude that a path $(v_0, v_1), \ldots, (v_{l-1}, v_{v_l})$ between nodes v_0 and v_l necessarily exists[1]. This transitivity has the important mathematical consequence that the entries in the k-th power A^k of the adjacency matrix A of a static network topology count all possible paths of length k between all possible pairs of nodes. Furthermore, transitivity of paths is the basis for a wealth of *algebraic network-analytic methods* such as spectral partitioning, the analysis of dynamical processes based on eigenvectors and eigenvalues, or the computation of centrality measures that are based on eigenvalue problems.

Notably, the property of transitivity of paths in static networks does *not* extend to time-respecting paths in temporal networks. Here, two time-respecting paths $(v_0, v_1; t_1), \ldots, (v_{k-1}, v_k; t_k)$ and $(v_k, v_{k+1}; t_{k+1}), \ldots, (v_{l-1}, v_l; t_l)$ only translate into a time-respecting path between v_0 and v_l if $t_k < t_{k+1}$ and, assuming that we impose a maximum time difference δ, if $0 < t_{k+1} - t_k \leq \delta$.

The simple observation that transitivity of paths holds in static networks, while it does not necessarily hold in temporal networks implies that by an analysis of static, time-aggregated

[1]Note though that this transitive path may or may not be the shortest path between the two nodes.

networks, we may overestimate transitivity in temporal networks. We can again illustrate this using our simple example of Figure 3.1, which shows two temporal networks G_1 and G_2 that are both consistent with the same (weighted) time-aggregated network shown in Figure 3.1c. Here, judging from the presence of a path $(a, c), (c, d)$ in the time-aggregated network, we may think that a time-respecting path connecting node a to d exists in the underlying temporal network. Looking at the two temporal networks G_1 and G_2 shown in Figure 3.1a and Figure 3.1b respectively, we see that at least for small values for the maximum time difference δ (such as $\delta = 1$) a corresponding time-respecting path only exists in the temporal network G_1, while it is absent in G_2.

3.3 Higher-order aggregate networks

In the previous section we have seen that for large maximum time differences δ we expect the shortest time-respecting paths to be rather similar to the shortest path in a static, time-aggregated representation. This is an intuitive result since by using large maximum time differences δ, we apply an implicit "aggregation" of time stamps which may nevertheless be far apart in the temporal dimension. At the same time, we observe that for small values of δ the temporal characteristics of the network result in time-respecting path structures that are markedly different from those in the static, time-aggregated network. As argued above, this implies that dynamical processes which evolve at time scales similar to that of the temporal network will be significantly affected by these path structures. It further questions the usefulness of path-based centrality measures that are computed based on the commonly used time-aggregated representation of temporal networks.

In this section, we introduce *higher-order time-aggregated networks*, a simple yet powerful abstraction of temporal networks which can be used to address some of the aforementioned problems. It can be seen as a simple generalization of the usual *first-order* time-aggregated representation introduced in Section 2.3.2.

3.3.1 *k*-th order aggregate networks

The key idea behind this abstraction is that the commonly used time-aggregated network is the simplest possible time-aggregated representation whose weighted links captures the frequencies of time-stamped links. Considering that each time-stamped link is a time-respecting path of length one, it is easy to generalize this abstraction to *higher-order time-aggregate networks* in which weighted links capture the frequencies of longer time-respecting paths. For a temporal network $G^T = (V, E^T)$ we thus formally define a k-th

order time-aggregated (or simply aggregate) network as a tuple $G^{(k)} = (V^{(k)}, E^{(k)})$ where $V^{(k)} \subseteq V^k$ is a set of node k-tuples and $E^{(k)} \subseteq V^{(k)} \times V^{(k)}$ is a set of links. For simplicity, we call each of the k-tuples $v = v_1 - v_2 - \ldots - v_k$ ($v \in V^{(k)}, v_i \in V$) a *k-th order node*, while each link $e \in E^{(k)}$ is called a *k-th order link*. We further assume that a k-th order link (v, w) between two k-th order nodes $v = v_1 - v_2 - \ldots - v_k$ and $w = w_1 - w_2 - \ldots - w_k$ exists if they overlap in exactly $k-1$ elements such that $v_{i+1} = w_i$ for $i = 1, \ldots, k-1$. The basic idea behind this construction is that each k-th order link (v, w) represents a possible time-respecting path of length k in the underlying temporal network, which connects node v_1 to node w_k via k time-stamped links Resembling so-called *De Bruijn graphs*[22], the basic idea behind this construction is that each k-th order link (v, w) represents a possible time-respecting path of length k in the underlying temporal network, which connects node v_1 to node w_k via k time-stamped links

$$(v_1, v_2 = w_1; t_1), \ldots, (v_k = w_{k-1}, w_k; t_k) \tag{3.1}$$

In analogy to the weights in a usual (first-order) aggregate representation, we further define the weights of such k-th order links by the frequency of the underlying time-respecting paths in the temporal network. Considering a maximum time difference δ and two k-th order nodes $v = v_1 - v_2 - \ldots - v_k$ and $w = w_1 - w_2 - \ldots - w_k$ we thus define

$$\omega(v, w) := |P(v, w, \delta)| \tag{3.2}$$

where

$$P = \{(v_1, v_2 = w_1; t_1), \ldots, (v_k = w_{k-1}, w_k; t_k) : 0 < t_{i+1} - t_i \leq \delta\} \tag{3.3}$$

is the set of all time-respecting paths in the temporal network that i) consist of the sequence of links indicated in Eq. 3.1, and ii) are consistent with a given maximum time difference of δ.

The higher-order aggregate network construction introduced above has a number of advantages. First and foremost, it provides a simple static abstraction of a temporal network which can be studied by means of standard methods from (static) network analysis. Each static path of length l in a k-th order aggregate network can be mapped to a time-respecting path of length $k+l-1$ in the original network. Importantly, and different from a first-order representation, k-th order aggregate networks allow to capture *non-Markovian characteristics* of temporal networks. In particular, they allow to represent temporal networks in which the k-th time-stamped link $(v_k = w_{k-1}, w_k)$ on a time-respecting path depends on the $k-1$ previous time-stamped links on this path. With this, we obtain a simple static network topology that contains information both on the presence of time-stamped links

in the underlying temporal network, as well as on the *ordering* in which sequences of k of these time-stamped links occur.

3.3.2 Second-order aggregate networks

In the following, we illustrate our approach by constructing second-order aggregate representations of the two temporal networks G_1 and G_2 shown in Figure 3.1. Both G_1 and G_2 are consistent with the same first-order time-aggregated network. We can easily generate second-order time-aggregated networks of the two temporal networks by extracting all time-respecting paths of length two (and assuming a given maximum time difference δ). For simplicity, in the following we limit our study to $\delta = 1$. For the temporal network G_1 shown in Figure 3.1a, we observe the following four different time-respecting paths of length two:

$$(a, c; 1), (c, e; 2)$$
$$(b, c; 4), (c, d; 5)$$
$$(b, c; 7), (c, e; 8)$$
$$(a, c; 10), (c, d; 11)$$

Based on the definition of links and link weights outlined above, we thus obtain the following four weighted second-order links:

$$\omega(a - c, c - e) = 1$$
$$\omega(b - c, c - d) = 1$$
$$\omega(b - c, c - e) = 1$$
$$\omega(a - c, c - d) = 1$$

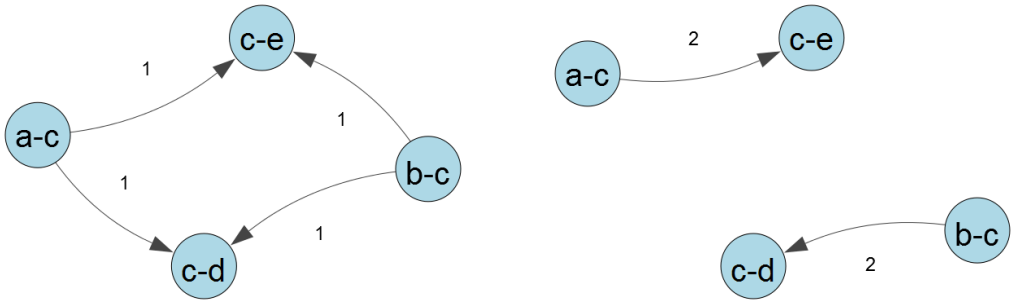
The resulting second-order network is depicted in Figure 3.2a Applying the same methodology to the temporal network G_2 shown in Figure 3.1b we obtain the following four time-respecting paths of length two

$$(a, c; 1), (c, e; 2)$$
$$(b, c; 4), (c, d; 5)$$
$$(b, c; 7), (c, d; 8)$$
$$(a, c; 10), (c, e; 11)$$

from which we obtain the following two weighted second-order links:

$$\omega(a-c, c-e) = 2$$
$$\omega(b-c, c-d) = 2$$

The resulting second-order aggregate network is shown in Figure 3.2b. Here we observe



(a) Temporal network G_1 (b) Temporal network G_2

Figure 3.2: Higher-order representation Second-order aggregate networks $G^{(2)}$ corresponding to the two temporal networks shown in Figure 3.1.

that, even though the two temporal networks G_1 and G_2 only differ in the order of two time-stamped links, the resulting second-order aggregate network is markedly different. The second-order network of G_1 indicates time-respecting paths connecting node a to both nodes e and d (both paths passing via node c). In particular, this corresponds to the connectivity that we would expect based on the transitivity of static paths in the first-order aggregate network shown in Figure 3.1c. The second-order network shown in Figure 3.2b reveals that the transitive path $(a, c), (c, d)$ in the first-order aggregate network does not translate to a time-respecting path in the temporal network G_2.

Clearly, the second-order aggregate networks illustrated above are only a special, particularly simple type of general, higher-order aggregate networks. Nevertheless, in the following chapters we will demonstrate that it contains important information about the causal topology of temporal networks which can help us in the analysis of temporal networks. In what follows, we will thus provide an in-depth study of second-order aggregate representations of six empirical data sets that will be introduced in the following section.

3.4 Data Sets

In our studies we investigate six empirical data sets representing different types of temporal networks. Some of them are time-stamped while others only consists of paths statistics. A summary of the data sets can be found in table 3.1. In the following we provide a brief description for each of the six data sets. Most of the data already comes in suitable format of time-stamped links,i.e. $(v, w; t)$, that directly can be transformed into a temporal network. In the subsequent sections we discuss how time-aggregate network can be constructed from the data and how a maximum time difference can be chosen.

Ants (AN) The ants data is the only system we consider that is not based on human interactions or movements. It covers time-stamped antenna-antenna interactions inferred from a filming of ants in an ant colony. The interactions are between individual pairs of ants that are part of the colony. We used the largest data set provided by Blonder and Dornhaus [17]. More specifically the first filming of colony 1 with a total of 1911 antenna-body interactions between 89 ants recorded over a period of 1438 seconds. The data comes already in time-stamped link format, e.g. *GBGW WBYG 1093* where the number indicate the time in terms of seconds when the interaction was recorded.

E-Mail (EM) This data set covers E-Mail exchanges inside a company [102]. The data was recorded over a period of nine months between 167 employees of a medium-size manufacturing company. We used a subset covering close to $11,000$ E-mail exchanges occurring during the first month of the observation period. The data comes in format of a link list from sender to recipient followed by the time-stamp e.g. *17;37;2010-01-02 17:40:10*. Messages with multiple recipients are recorded as separate links.

Hospital (HO) The hospital data consists of time-stamped contacts between 46 health-care workers and 29 patients in a hospital in Lyon [164]. Contacts have been recorded via proximity sensing badges in the week from Dec. 6 to Dec. 10 2010. For our analysis, we used a subset of more than $15,000$ contacts occurring within the first 48 hours of the observation period. The data comes with a time-stamp followed by the id's of the interacting people. Additionally, the data also contains the profession of the person that we don't consider in our study, e.g. *6240 1152 1193 MED NUR*.

Reality Mining (RM) Another human interaction network we consider is based on the reality mining project, a contact network of students and academic staff members at

a university campus. The data uses time-stamped proximity data recorded via Bluetooth-enabled phones at a university campus over a period of more than six months [34]. In our study we extracted a subset covering the week from September 8th to 15th 2004, which comprises a total of 26 260 time-stamped interactions between 64 individuals. If two individuals are in the range of 5-10m to each other it is recorded as interaction. The data provides a list like format of the corresponding MAC address of phones in close enough range from which we extract the interactions as links, e.g. *61961025059 61961024891 21.*

Flights (FL) The airline transportation networks represents multi-segment itineraries of airline passengers on domestic flights in the United States. The data has been extracted from the freely available RITA TranStats Airline Origin and Destination Survey (DB1B) database [2], which contains 10 % samples of all airline tickets sold in the United States for each quarter since 1993. For our study we extracted 230 000 multi-segment flights ticketed by American Airlines in the fourth quarter of 2001, which connect a total of 116 airports in the United States. There are no time-stamps in this data set but only pathways of flight tickets. Hence it is not a "real" temporal network but rather a topological one. For each ticket number i, an itinerary consists of a time-ordered sequence of multiple flight segments between airports indicated by their three-letter IATA code. An example for a time-ordered itinerary with ticket number i is given in the following:

$$
\begin{aligned}
&i, CLT, ORF \\
&i, ORF, LGA \\
&i, LGA, ORF \\
&i, ORF, CLT
\end{aligned}
$$

While no precise time stamps are known for individual segments, their ordering allows to directly construct time-respecting flight paths taken by individual passengers. For the example above, a time-respecting path

$$(CLT, ORF; 1) \rightarrow (ORF, LGA; 2) \rightarrow (LGA, ORF; 3) \rightarrow (ORF, CLT; 4)$$

can be constructed. Here time-respecting paths necessarily consist of interactions which *immediately follow each other in subsequent time steps*, since otherwise a spurious flight path $(CLT, ORF; 1) \rightarrow (ORF, CLT; 4)$ would be inferred for the example above. Furthermore, a time-respecting path is only inferred if the ticket number of consecutive flight segments is identical.

London Tube (LT) This data set captures passenger journeys in the London underground transportation network. The data has been extracted from the freely available Rolling Origin & Destination Survey (RODS) database [92] provided by the London Underground. The RODS database covers a 5 % sample of all journeys made by passengers who used the *Oyster* electronic ticketing card during a period of one week. This amounts to a total of more than four million passenger flows between 309 London Underground stations. The flow data comes without precise time-stamps but allows a topological construction of the passenger movements. A flow from A to B contains of the total of all passengers that entered the London Underground at station A and left it at station B, i.e. *Acton Town;Alperton;325*. To get the full path from A to B we map those passenger flows to a network representation of the London Underground. The shortest path between A and B on the network provides us with detailed itinerary data. We then computed time-respecting paths based on directly consecutive travel segments like $A \to S_1 \to S_2 \to \dots B$.

3.4.1 Higher-order time-aggregated networks from data

Extracting time-respecting paths in the six data sets allows us to construct higher-order time-aggregated representations of the underlying temporal interaction sequences. We apply our approach to the the six data sets described in section 3.4. A temporal network G^T consists of directed time-stamped links $(v, w; t)$ for nodes v and w and discrete time stamps t. A *first-order time-aggregated network* $G^{(1)}$ can then be defined, where a directed link (v, w) between nodes v and w exists whenever a time-stamped link $(v, w; t)$ exists in G^T for some time stamp t. In addition, link weights $w^{(1)}(v, w)$ can be defined as the (relative) number of link occurrences in the temporal network. Considering that links can be thought of as time-respecting paths of length one, we can similarly construct a *second-order time-aggregated network* by considering time-respecting paths of length two. For this, we define a second-order time-aggregated network $G^{(2)}$ as tuple $(V^{(2)}, E^{(2)})$ consisting of *second-order nodes* $V^{(2)}$ and *second-order links* $E^{(2)}$. Second-order nodes $e \in V^{(2)}$ represent links in the first-order aggregate network $G^{(1)}$. Second-order links $E^{(2)}$ represent all possible time-respecting paths of length two in $G^{(1)}$. Based on the definition of time-respecting paths with a maximum time difference δ, *second-order link weights* $w^{(2)}(e_1, e_2)$ can be defined based on the frequency of two-paths, i.e. the frequency of time-respecting paths $(a, b; t_1) \to (b, c; t_2)$ of length two in G^T (for $1 \leq t_2 - t_1 \leq \delta$). Since multiple two-paths $(a', b; t) \to (b, c'; t')$ can pass through node b at the same time, it is necessary to proportionally correct second-order link weights for all multiple occurrences.

For the simple case $\delta = 1$, one can define second-order link weights as

$$w^{(2)}(e_1, e_2) := \sum_t \frac{\delta_{(a,b;t-1)}\delta_{(b,c;t)}}{\sum_{a',c' \in V} \delta_{(a',b;t-1)}\delta_{(b,c';t)}}, \tag{3.4}$$

where $\delta_{(a,b;t)} = 1$ if linke $(a, b; t)$ exists in the temporal network G^T and $\delta_{(a,b;t)} = 0$ otherwise. Following the arguments above, it is simple to generalize weights to capture two-paths $(a, b; t_1) - (b, c; t_2)$ for $1 \leq t_2 - t_1 \leq \delta$.

3.4.2 Choice of a maximum time difference

Regarding the choice of a reasonable maximum time difference δ for the notion of shortest time-respecting paths as discussed in section 3.2.1, we emphasize that the choice of this parameter needs to be adapted to the inherent time scale of the network evolution in each of the six data sets individually. In general, such a choice is non-trivial as it heavily influences i) whether or not pairs of nodes can reach each other, and ii) to what extent temporal characteristics influence the structures of time-respecting paths. In particular, for too small choices of δ the definition of time-respecting paths is likely to be too restrictive and almost no paths will be found [69, 119]. Contrariwise, the choice of a too large value for δ results in the fact that we effectively "aggregate" the time-stamped sequence of links, thus discarding information about the detailed ordering and timing of links. In general, we have chosen the maximum time difference δ as the smallest possible value such that the set of nodes that can mutually influence each other via time-respecting paths (i.e. the strongly connected component) represents a sizable fraction of the network. For the (AN) data, a maximum time difference δ of six seconds was applied, which gives rise to a subset of 61 nodes that can reach each other via time-respecting paths. For the (EM) data set, we used a maximum time difference δ of 60 minutes, resulting in a subset of 96 employees mutually connected via time-respecting paths. For the (HO) and (RM) data sets, we used a maximum time difference δ of five minutes, which resulted in a subset of 53 and 58 individuals respectively who can mutually reach each other via time-respecting paths. While the condition of maximum time difference δ is crucial for the (AN), (RM), (HO) and (EM) data sets, for the (FL) and the (LT) data set we can use a strict definition of a time-respecting path and only consider consecutive travel segments. For the (FL) data set, the strongly connected component comprises 116 airports, while it comprises 132 underground stations in the (LT) data set.

In general, the optimum choice of δ depends on i) temporal characteristics of the temporal network under investigation, and ii) the time scale of the dynamical processes one is

Data Sets	Nodes	Links	δ	Type
Ants (AN)	61	300	6 sec	Animal interaction
E-Mail (EM)	96	701	60 min	Messaging
Hospital (HO)	53	928	5 min	Human interaction
RealityMining (RM)	58	899	5 min	Human interaction
Flights (FL)	116	1316	itineraries	Transportation
London Tube (LT)	132	334	itineraries	Transportation

Table 3.1: Summary of the six data sets used for the temporal networks study While AN, EM, HO and RM are based on time-stamped data, FL and LT are merely based on itineraries.

interested in. In this study, we have used a rather simple and heuristic method, defining δ as the smallest value that still renders the system strongly connected in a temporal sense. In general a more principled approach which, e.g., inherently couples the choice of δ to a characterization of inter-event time statistics would be desirable.

3.5 Conclusion

In this chapter we introduced the framework of higher-order aggregate networks. This is an abstraction of temporal networks that relies on the statistics of time-respecting paths. The concept of time-respecting paths requires time-stamped links and the fixation of a maximum time-difference δ. All links that happen within the period δ are considered to be consecutive hence allowing the formation of a network path. Higher-order aggregate network explicitly account for the causality present in time-respecting paths up to a given length k. Therefore higher-order aggregate networks include additional information about the ordering of links that is not possible in terms of time-aggregated first-order networks

Further we have shown how we can extract higher-order aggregate networks from real-world data sets. Since, the framework explicitly focus on path structures we can also analyze data sets where no time-stamps are available but we instead have a given ordering of links. In the following two chapters we show how the presented framework can be used to approach the analysis of properties of temporal systems and how it compares to the classical time-aggregated perspective.

Chapter 4

Temporal Centralities

Summary

Addressing temporal ordering, we introduce a novel framework for the study of *path-based centralities* in temporal networks. Focusing on betweenness, closeness and reach centrality, we first show that an application of these measures to time-aggregated, static representations of temporal networks yields misleading results about the actual importance of nodes. To overcome this problem, we define path-based centralities in *higher-order aggregate networks*, a previously introduced generalization of the commonly used static representation of time-stamped data. Using data on six empirical temporal networks, we show that the resulting higher-order measures better capture the true, *temporal* centralities of nodes. This results highlight that higher-order aggregate networks are a powerful tool to incorporate path-based statistics in temporal networks.

Based on Scholtes, I., Wider, N. and Garas. A. *Higher-Order Aggregate Networks in the Analysis of Temporal Networks: Path structures and centralities* , Eur. Phys. J. B 89 (3) 61, 2016. NW contributed to the concept of the temporal centrality measures, implemented the reach centrality, analyzed the data and discussed the results. NW also contributed to the development of the higher-order aggregate networks.

4.1 Introduction

Understanding and analyzing complex systems bears a variety of challenges. Interaction between elements of a systems not only influence the directly involved actors but can also lead to emergent effects on the systemic level. To investigate such dynamics it is not only important to know the immediate interactions of an individual but also its *importance* in respect to the whole system. In network theory this question is addressed with *centrality* measures.

Some of the first studies that addressed the centrality of individuals originated from social science. It was investigated how central positions and leadership increases the group efficiency in solving problems [12, 13, 147]. Different communication patterns between group members were analyzed and it was found that better connected groups increase their performance. Further, leaders among initially equal group members often emerged in *central* positions of the communication. All lot of similar studies and experiments followed to entangled the effects of different centrality measures. The results varied across different research leading to sometimes confusing and contradicting conclusions [24]. The concept of centrality is tied to the contextual application and therefore has to be interpreted carefully.

What kind of positions inside a system are regarded central varies according to its definition. In a social network a person with a lot of friends may be in a good position to spread his ideas. On the other hand a lot of social contacts makes the person also more prone to get influenced. Also in technical [64, 172], economical [128] and financial [11] systems centrality is an important issue. Banks that have a lot of liabilities are at higher risk during a financial crisis. In power grids it is important to know how to prevent blackouts by securing or supporting crucial connectors in the network. In trade networks not necessarily the amount of trade relations is most important but rather a unique position in connecting different parts of the systems such that distant trade relationship are channeled through the one node. The context is important in all this examples to asses what kind of centrality is relevant for the intended study.

The framework of centrality measures plays an important role in network analysis. New and refined measures are developed all the time to keep up with specific needs. Even though, more elaborate tools get available this measures are normally only applied to static networks. Temporal data is usually aggregated over certain time-windows before it analyzed with casual measures. However, the temporal structure of time-respecting path gets destroyed in this process. A static centrality analysis can lead to misleading results about the importance of nodes in respect to time-respecting processes or topologies [114, 155, 157]. Centralities that are based on paths connecting distant nodes are directly

affected.

Highlighting the important consequences introduced by the specific ordering of links in real-world temporal networks, in this chapter we study how this ordering affects path-based centrality measures in temporal networks. Building on the concept of time-respecting paths with a maximum time difference between consecutive links as previously discussed in Chapter 3 and Refs. [69, 119], we introduce three different notions of path-based temporal node centralities which emphasize the additional *temporal-topological* dimension that is introduced due to the ordering of links in temporal networks. We generalize usual static network centralities and take into account the temporal ordering that preserves time-respecting paths. In particular, we formally define temporal variations of *betweenness*, *closeness* and *reach* centrality and demonstrate how they can be computed based on the topology of shortest time-respecting paths emerging in temporal networks. Calculating these temporal centrality measures for six empirical data sets, we quantify to what extent a ranking of nodes based on temporal centralities coincides with a ranking of nodes based on the same measures, however calculated based on the corresponding static, time-aggregated networks. From our results we conclude that a static analysis of node centralities yields misleading results about the importance of nodes with respect to time-respecting paths.

Generalizing the usual time-aggregated static perspective on temporal networks, we further develop the second-order time-aggregated representations introduced in Chapter 3, obtaining higher-order time-aggregated representations which can be conveniently analyzed using standard network-analytic methods. Notably, despite being static representations of temporal networks, we show that these higher-order representations allow to incorporate those order correlations that have been shown to influence the causal topologies of temporal networks. We finally define generalizations of static betweenness, closeness and reach centrality based on a second-order aggregate representation of temporal networks. Using six data sets on temporal networks, we show that these second-order generalizations of centralities constitute highly accurate approximations for the centrality of nodes calculated based on the detailed time-respecting path structures in temporal networks.

The remainder of this chapter is structured as follows: In Section 4.2 we define three temporal centrality measure which account for the temporal-topological characteristics introduced by the shortest time-respecting path structures in real-world temporal networks. Comparing the importance of nodes according to i) temporal centralities, ii) centralities calculated based on a commonly used static, time-aggregated representation, and iii) second-order centralities calculated based on a static, second-order time-aggregated representation, we show that higher-order aggregate networks provide interesting perspectives for the analysis of temporal networks.

4.2 Temporal node centralities

We will focus our analysis on three widely adopted path-based notions of centrality, namely i) betweenness, ii) closeness and iii) reach centrality. The rationale behind this choice is that all of these three measures can easily be computed based on paths in time-aggregated networks, while they additionally facilitate a straight-forward extension to temporal networks based on the notion of shortest time-respecting paths (c.f. similar extensions studied in [69, 81, 154]). In the following, we first formally define the *temporal* betweenness, closeness and reach centrality of nodes. We then compute the resulting measures for all nodes based on the actual shortest time-respecting paths in the time-stamped link sequences in our six data sets (and using the individually determined maximum time difference δ). The resulting centrality scores are considered as the *ground-truth* against which we then compare the centrality scores resulting from the application of the same centrality measures to i) the commonly used (first-order) time-aggregated representation, and ii) a second-order aggregate network representation of the corresponding temporal network.

4.2.1 Temporal betweenness centrality

We first address the question to what extent the temporal betweenness centrality of nodes in a temporal network can be approximated by means of static betweenness centralities calculated based on static, time-aggregated representations. To this end, we first formally define the temporal betweenness centrality of a node in a temporal network. According to the common definition, the (unnormalized) betweenness centrality of a node v is simply calculated as the total number of shortest paths passing through node v [40]. Highlighting the fact that we can directly apply this measure to first-order time-aggregated networks, we thus define the *first-order betweenness centrality* $\mathrm{BC}^{(1)}(v)$ of a node v as

$$\mathrm{BC}^{(1)}(v) := \sum_{u \neq v \neq w} |P^{(1)}(u, w; v)| \tag{4.1}$$

where $P^{(1)}(u, w; v)$ denotes the set of those shortest paths from node u to w in a static network that pass through node v.

Applying this idea to temporal networks, a straight-forward way to define the *temporal betweenness centrality* of a node is to count all shortest *time-respecting paths* passing through it. However, and as mentioned in Section 3.2, temporal networks introduce the complication that, in order to unambiguously define shortest time-respecting paths, we need to include a start time t_0 starting from which time-respecting paths are to be considered. For

each pair of nodes u, v and each start time t_0 we can thus directly define an instantaneous distance function for a temporal network as

$$\text{dist}^{\text{temp}}(u, v, t_0) := \text{len}(p), p \in P^{\text{temp}}(u, v, t_0) \tag{4.2}$$

where $P^{\text{temp}}(u, v, t_0)$ denotes the set of shortest
time-respecting paths from u to v that start at time t_0 (and which are consistent with a given maximum time difference δ). Based on this instantaneous definition of shortest time-respecting paths, we can further define a distance function that gives the minimum distance across *any* start time as follows:

$$\text{dist}^{\text{temp}}(u, v) := \min_{t_0} \text{dist}^{\text{temp}}(u, v, t_0) \tag{4.3}$$

With this we can further define the set of shortest time-respecting paths across all start times as

$$P^{\text{temp}}(u, v) := \bigcup_{t_0} \{p \in P^{\text{temp}}(u, v, t_0) | \text{len}(p) = \text{dist}^{\text{temp}}(u, v)\} \tag{4.4}$$

i.e. we only consider those (instantaneous) shortest time-respecting paths whose lengths correspond to the minimum shortest time-respecting length across *all* possible start times. We can now define the *temporal betweenness centrality* $\text{BC}^{\text{temp}}(v)$ of a node v in analogy to Eq. 4.1 as

$$\text{BC}^{\text{temp}}(v) := \sum_{u \neq v \neq w} |P^{\text{temp}}(u, w; v)| \tag{4.5}$$

where $P^{\text{temp}}(u, w; v)$ denotes the set of those shortest time-respecting paths across all start times which connect node u to w and which pass through node v.

Let us illustrate this definition using the temporal networks shown in Figure 3.1a and Figure 3.1b. Applying the static betweenness centrality as defined in Eq. 4.1 to the first-order aggregate network shown in Figure 3.1c, we find that for node c we have $\text{BC}^{(1)}(c) = 4$, while for all other nodes we have a betweenness centrality of zero. Again assuming $\delta = 1$, for the temporal betweenness centrality of node c in network G_1 shown in Figure 3.1a, we find that indeed four shortest time-respecting paths pass through node c, i.e. we have $\text{BC}^{\text{temp}}(c) = 4$ while we again have a zero temporal betweenness centrality for all other nodes. Notably, in this particular case the temporal betweenness centralities of nodes correspond to the betweenness centralities of nodes calculated based on the first-order time-aggregated network. This happens because all paths in the first-order aggregate

network have a counterpart in terms of a shortest time-respecting path.

However, in Section 3.2 we have seen that, in general, shortest time-respecting paths in temporal networks may not coincide with shortest paths in the (first-order) time-aggregated network. As a consequence, the temporal betweenness centralities of nodes may differ from the first-order betweenness centralities calculated from a static, first-order aggregate representation. This can be seen for the temporal network G_2 shown in Figure 3.1b. Based on the temporal sequence of time-stamped links, here we find only two different shortest time-respecting paths passing through node c, namely one connecting node a via c to e and a second one connecting node b via c to d. The two additional shortest time-respecting paths found in G_1 are absent in G_2, therefore in G_2 node c has a temporal betweenness centrality $\mathrm{BC}^{\mathrm{temp}}(c) = 2$, thus being, at least from the perspective of temporal betweenness centrality, less important than in G_1.

In the following we study the question to what extent first-order betweenness centralities can be used as a proxy for the temporal betweenness centralities of nodes in our six data sets of real-world temporal networks introduced in Section 3.4. In particular, we study this question in the following way: For each node v in the six data sets we calculate i) the first-order betweenness centrality $\mathrm{BC}^{(1)}(v)$ based on the first-order aggregate network, as well as ii) the (ground truth) temporal betweenness centrality $\mathrm{BC}^{\mathrm{temp}}(v)$ based on actual shortest time-respecting paths in the temporal network. We then assess the correlation between both measures by computing the Pearson correlation coefficient (as well as the corresponding p-value) for the sequence of paired values $(\mathrm{BC}^{(1)}(i), \mathrm{BC}^{\mathrm{temp}}(i))$ for all nodes $i \in V$.

Since centrality scores of nodes in networks are often used and interpreted in a relative fashion, we further perform an additional analysis that accounts for variations in the actual centrality values, which however may not affect the relative importance of nodes. For this, we first rank nodes according to their temporal and first-order betweenness centralities respectively. We then calculate the Kendall-Tau rank correlation coefficient in order to quantitatively assess to what extent nodes are ranked similarly according to both notions of centrality (even though the actual centrality values for these nodes may differ).

The results of this analysis are shown in the left column of Table 4.1, in which we report both the Pearson as well as the Kendall-Tau rank correlation coefficients between the temporal and the first-order betweenness centralities of nodes for each of the six data sets introduced before. Here, a first interesting result is that both the Pearson and the Kendall-Tau rank correlation coefficients exhibit a large variation between 0.80 and 0.99, as well as 0.62 and 0.81 respectively. The results indicate that, depending on the characteristics of the underlying temporal network, temporal betweenness centralities can be reasonably

	$BC^{temp} \sim BC^{(1)}$		$BC^{temp} \sim BC^{(2)}$	
	Pearson	Kendall-Tau	Pearson	Kendall-Tau
Ants (AN)	0.82 (3.49e-16)	0.64 (2.05e-13)	0.80 (1.96e-14)	0.59 (1.94e-11)
E-Mail (EM)	0.80 (3.29e-22)	0.73 (8.36e-26)	0.97 (7.52e-60)	0.74 (1.11e-26)
Hospital (HO)	0.93 (2.39e-23)	0.81 (1.18e-17)	0.96 (2.36e-30)	0.87 (5.55e-20)
RealityMining (RM)	0.95 (2.83e-30)	0.62 (7.28e-12)	0.93 (3.74e-26)	0.75 (1.12e-16)
Flights (FL)	0.99 (6.91e-108)	0.66 (9.09e-26)	0.99 (2.66e-98)	0.65 (4.25e-25)
London Tube (LT)	0.85 (2.58e-37)	0.66 (1.22e-29)	0.87 (3.28e-42)	0.71 (9.32e-34)

Table 4.1: Pearson and Kendall-Tau rank correlation coefficients between temporal betweenness centrality (ground truth) and betweenness centrality calculated based on the first-order aggregate network and the second-order aggregate network. Values in parentheses indicate the p-value.

well approximated by first-order betweenness centrality for some data sets (e.g., for (FL), (HO), (RM)) while such an approximation should be taken with caution for other data sets.

Based on these results it is reasonable to ask if we can better approximate temporal centrality, especially for those data sets where the correlation between the first-order and the temporal betweenness centrality is comparably weak. In Section 3.3 we have argued that the generalization of higher-order aggregate networks allows to construct static representations of temporal networks that capture both temporal and topological characteristics that emerge from the ordering of links and the statistics of time-respecting paths. Focusing on a second-order representation, in the remainder of this section we will study to what extent second-order aggregate networks can be used in the analysis of temporal node centralities.

Importantly, such an analysis is facilitated by the fact that second-order aggregate networks are *static networks*, which allows for a straight-forward application of standard centrality measures to the second-order topology. In the case of second-order aggregate networks, applying standard centrality measures we obtain centrality values for higher-order nodes (v, w), each of the higher-order nodes being a k-tuple of nodes in the first-order network. In order to arrive at a centrality measure for the original (first-order) nodes, we thus must project this measure to the level of nodes in the first-order network.

Luckily, this can be done in a simple way which we outline in the following: For a second-order network $G^{(2)} = \left(V^{(2)}, E^{(2)}\right)$, let us first define a second-order distance function $\text{dist}^{(2)}(v, w)$ which, for each pair of *first-order* nodes $v, w \in V^{(1)}$, gives the length of a

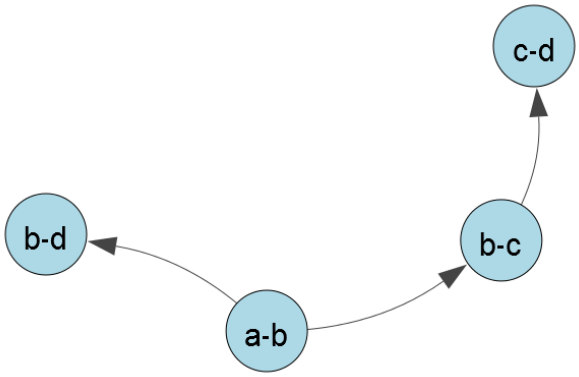


Figure 4.1: Simple example for a second-order aggregate network

shortest path based on the topology of the *second-order* aggregate network as

$$\mathrm{dist}^{(2)}(v,w) := \min_{\substack{x,y \in V^{(2)} \\ x=v-* \\ y=*-w}} L^{(2)}(x,y) + 1 \tag{4.6}$$

where $L^{(2)}(x,y)$ denotes the length of a shortest path between the second-order nodes $x, y \in V^{(2)}$. The rationale behind this definition is that in the second-order aggregate network, we can have multiple shortest paths with different lengths between different second-order nodes, which nevertheless map to paths between a single pair of first-order nodes. As an example, consider the two first-order nodes a and d in the simple second-order network shown in Figure 4.1. Here we observe that, from the perspective of second-order nodes, both $(a-b, b-d)$ as well as $(a-b, b-c), (b-c, c-d)$ are shortest paths (between different pairs of nodes) in the second-order network with lengths $L^{(2)}(a-b, b-d) = 1$ and $L^{(2)}(a-b, c-d) = 2$ respectively. However, from the perspective of first-order nodes both of these second-order paths connect node a to node d (via paths of length 2 and 3 respectively). Using the definition from Eq. 4.6 thus allows us to correctly calculate the second-order distance between a and d as $\mathrm{dist}^{(2)}(a,d) = L^{(2)}(a-b, b-d) + 1 = 2$.

The above definition of a second-order distance function now allows us to define a *second-order betweenness centrality* $\mathrm{BC}^{(2)}(v)$ of a node v based on Eq. 4.1. For this, we simply count all second-order shortest paths between two nodes u and w which i) pass through node v, and ii) whose length corresponds to the second-order distance $\mathrm{dist}^{(2)}(u,v)$. For-

mally, we define

$$\mathrm{BC}^{(2)}(v) := \sum_{\substack{x \neq y \in V \\ u-x \in V^{(2)} \\ y-w \in V^{(2)}}} |\{p \in P^{(2)}(u-x, y-w; v) : \mathrm{len}(p) = \mathrm{dist}^{(2)}(u, w)\}| \tag{4.7}$$

where, in analogy to $P^{(1)}(u, w; v)$ above, $P^{(2)}(u-x, y-w; v)$ denotes the set of all shortest paths in the second-order network that connect node $u-x$ to $y-w$ and that pass through a first-order node v.

With this, we have defined a second-order betweenness centrality which allows to calculate node centralities in a way that incorporates the causal topology as captured by the second-order aggregate network. Let us again illustrate this approach using the simple examples shown in Figure 3.1. For the temporal network G_1 we can compute a second-order betweenness centrality based on the second-order network shown in Figure 3.2a. Here we observe a total of four shortest paths between pairs of nodes in the second-order network, namely:

$$(a-c, c-e)$$
$$(a-c, c-d)$$
$$(b-c, c-d)$$
$$(b-c, c-e)$$

For each node in the first-order network, we can now count the number of second-order shortest paths that they are on, obtaining $B^{(2)}(c) = 4$ while $B^{(2)}(x) = 0$ for all nodes $x \neq c$. In this particular case, the second-order betweenness centrality values exactly correspond both to the temporal as well as the first-order betweenness centralities. Again, this is different for the temporal network G_2 shown in Figure 3.1b. Considering the second-order aggregate network shown in Figure 3.2b, we only find the following two shortest paths in the second-order aggregate network

$$(b-c, c-d)$$
$$(a-c, c-e)$$

thus obtaining $\mathrm{BC}^{(2)}(c) = 2$. Here, we find that while the second-order betweenness centralities in G_2 corresponds to the temporal betweenness centralities, they differ from those calculated from the first-order aggregate network. The reason for this is that in the example G_2 shortest time-respecting paths of length two differ from what we would expect

based on the first-order network.

We emphasize that the exact correspondence between the second-order and the temporal betweenness centralities in the examples discussed above is because we have no shortest time-respecting paths of length three or longer, whose presence could differ from what we expect based on the second-order network. To what extent this affects the applicability of second-order aggregate networks in real-world scenarios is not clear and thus requires a further investigation. In the following, we thus study to what extent second-order betweenness centrality can be used to approximate the temporal betweenness centralities of nodes in the six real-world data sets studied above. For this, we first construct a second-order aggregate network as introduced in Section 3.3. We then calculate the betweenness centrality values $\mathrm{BC}^{(2)}(v)$ of all nodes v as described above, comparing the resulting centralities with the (ground-truth) temporal betweenness centralities $\mathrm{BC}^{\mathrm{temp}}(v)$.

The results of this analysis are shown in the right column of Table 4.1. Here we find that for most of the data sets, second-order betweenness centralities are correlated with the true, temporal betweenness centralities in a stronger way than the corresponding first-order approximation of betweenness centrality. For the (EM) data sets capturing E-Mail exchanges between employees in a manufacturing company, we observe an increase of the Pearson correlation coefficient ρ from 0.80 to 0.97, while the associated Kendall-Tau rank correlation coefficient τ increases rather mildly from 0.73 to 0.74. We attribute this to the fact that the second-order aggregate network better captures the structures of time-respecting paths in the temporal network compared to the first-order network. For the two data sets (HO) and (LT) we observe a similar increase both in the Pearson and the Kendall-Tau rank correlation coefficients, while the values remain largely unchanged for the (FL) data set. In particular, for the latter data set the first-order betweenness centrality already exhibits a correlation coefficient of 0.99 which indicates that in this particular case temporal characteristics do not significantly alter the structure of shortest time-respecting paths. For the two data sets (AN) and (RM) we observe a small decrease in the Pearson correlation values for the second-order approximation. Notably, for (RM) the decrease from 0.97 to 0.95 is accompanied by an increase of the Kendall-Tau coefficient from 0.62 to 0.75. This indicates that, even though the actual values of second-order betweenness centralities may be less correlated with temporal betweenness centralities than the first-order betweenness centralities, the second-order betweenness centralities provides us with a significantly better perspective on the *relative* importance of nodes.

Finally, for the (AN) data set we note that both the Pearson and the Kendall-Tau rank correlation coefficients are worse for the second-order betweenness centralities. While the interesting question in what respect the temporal characteristics of (AN) differ from those

of the other temporal networks remains to be investigated in more detail, we expect this result to be related to non-stationary properties. We particularly observe that some of the nodes (i.e. ants) are only active during certain phases of the observation period. This imposes a natural ordering of interactions which particularly prevents nodes which are only active during an early phase to be reachable from nodes which are only active at a later phase.

4.2.2 Temporal closeness centrality

Let us now turn our attention to *closeness centrality*, which captures a node's average distance to all other nodes in a network. For a directed, static (first-order aggregate) network the closeness centrality of a node v is commonly defined as

$$\mathrm{ClC}^{(1)}(v) = \sum_{u \neq v} \frac{1}{\mathrm{dist}^{(1)}(u, v)} \tag{4.8}$$

where the distance function $\mathrm{dist}^{(1)}(u, v)$ denotes the distance, i.e. the length of a shortest path, from node u to v in the first-order aggregate network.

We can easily define a temporal version of closeness centrality based on the temporal distance function $\mathrm{dist}^{\mathrm{temp}}(u, v)$ which we have defined in Eq. 4.3 in the context of temporal betweenness centrality. Here, we remind the reader that the function $\mathrm{dist}^{\mathrm{temp}}(u, v)$ captures the minimum length of a shortest time-respecting path across all possible start times t_0. Using this temporal distance function, we can apply the standard definition in Eq. 4.8 and define the *temporal closeness centrality* of a node v in a temporal network as

$$\mathrm{ClC}^{\mathrm{temp}}(v) = \sum_{u \neq v} \frac{1}{\mathrm{dist}^{\mathrm{temp}}(u, v)} \tag{4.9}$$

Let us again illustrate this definition using the temporal networks shown in Figure 3.1. Node e in the temporal network G_1 shown in Figure 3.1a can be reached from nodes a and b via two shortest time-respecting paths of length two, as well as from node c via a shortest time-respecting path of length one. For the temporal closeness centrality, we thus find $\mathrm{ClC}^{\mathrm{temp}}(e) = 2$. It is easy to confirm that this corresponds to the first-order closeness centrality of node e. Again a mere reordering of links can change the closeness centralities of nodes, as can be seen in the temporal network G_2 shown in Figure 3.1b. Here, we see that node e can only be reached from node a via a shortest time-respecting path of length two, as well as from node c via a shortest time-respecting path of length one. For node e

	$\mathrm{ClC}^{\mathrm{temp}} \sim \mathrm{ClC}^{(1)}$		$\mathrm{ClC}^{\mathrm{temp}} \sim \mathrm{ClC}^{(2)}$	
	Pearson	Kendall-Tau	Pearson	Kendall-Tau
Ants (AN)	0.91 (1.67e-24)	0.75 (1.54e-17)	0.96 (2.05e-35)	0.83 (4.80e-21)
E-Mail (EM)	0.93 (4.74e-44)	0.79 (4.96e-30)	0.98 (2.52e-71)	0.92 (1.54e-40)
Hospital (HO)	0.96 (2.09e-29)	0.83 (1.88e-18)	0.99 (1.46e-40)	0.90 (1.76e-21)
RealityMining (RM)	0.96 (1.03e-33)	0.77 (1.99e-17)	0.99 (1.64e-51)	0.89 (5.30e-17)
Flights (FL)	0.91 (3.35e-46)	0.81 (1.88e-18)	0.97 (4.57e-75)	0.93 (9.57e-50)
London Tube (LT)	0.98 (1.33e-91)	0.87 (2.57e-49)	0.98 (3.26e-92)	0.87 (1.07e-49)

Table 4.2: Pearson and Kendall-Tau rank correlation coefficients between temporal closeness centrality (ground truth) and closeness centrality calculated based on the first-order aggregate network and the second order aggregate network. Values in parentheses indicate the p-value.

in the temporal network G_2 we thus find a temporal closeness centrality $\mathrm{ClC}^{\mathrm{temp}}(e) = 1.5$, highlighting that it is, at least from the perspective of closeness centrality, less "important" than in the temporal network G_1.

Considering the example above we see that, due to the ordering and timing of links, first-order closeness centralities can be a misleading proxy for the temporal closeness centralities of nodes in temporal networks. In the following we thus again empirically study this question using our six data sets on temporal networks. We again use the temporal closeness centralities $\mathrm{ClC}^{\mathrm{temp}}(v)$ of nodes as the ground truth, then studying whether temporal closeness centralities can reasonably be approximated by first-order closeness centralities $\mathrm{ClC}^{(1)}(v)$. The results of this analysis are shown in the left column of Table 4.2, which reports the observed Pearson and Kendall-Tau rank correlation coefficients for each of the six data sets.

We observe again that the answer to the question of how well temporal closeness centralities can be approximated by first-order static closeness centralities depends on the actual data set. The lowest Pearson correlation coefficient of 0.91 is obtained for the (FL) and the (AN) data sets, while the highest Pearson correlation coefficient of 0.98 is obtained for (LT). The lowest Kendall-Tau rank correlation coefficient is 0.75 for (AN), while the highest value of 0.87 is achieved for (LT). We further observe that, compared to betweenness centralities, we generally obtain conceivably larger correlation values between temporal and first-order closeness centralities. This can intuitively be explained by the fact that, while temporal betweenness centralities are influenced by the actual *structure* of shortest time-respecting paths, temporal closeness centralities are merely influenced by their lengths. We thus expect temporal closeness centrality to be insensitive to characteristics of temporal networks that change the structure of paths but not their lengths, hence explaining the larger correlation coefficients.

Let us now study whether we can better approximate temporal closeness centralities using a generalization which is calculated based on the static, second-order aggregate representation of a temporal network. For this we first introduce how closeness centralities of nodes can be calculated based on a second-order aggregate network. We recall that in Eq. 4.6 we have defined a second-order distance function $\text{dist}^{(2)}(v, w)$ which provides us with the distance between (first-order) nodes based on shortest paths in a second-order aggregate network. This distance function allows us to directly define a *second-order closeness centrality* $\text{ClC}^{(2)}(v)$ as

$$\text{ClC}^{(2)}(v) = \sum_{u \neq v} \frac{1}{\text{dist}^{(2)}(u, v)} \tag{4.10}$$

i.e. for each node v in a network, we simply sum the inverse of the distances to all nodes according to the topology of the second-order aggregate network.

Again, we illustrate the notion of second-order closeness centrality using the two illustrative examples of temporal networks shown in Figure3.1. Figure 3.2a shows the second-order aggregate network corresponding to the temporal network G_1 shown in Figure 3.1a. Here we find that the second-order node $c - e$ can be reached via two shortest paths

$$(b - c), (c - e)$$
$$(a - c), (c - e)$$

of length one from the second-order nodes $b - c$ and $a - c$. Furthermore, we have an additional second-order "path" of length zero from node $c - e$ to itself. Using the second-order distance function as defined in Eq. 4.6, we thus infer the following values:

$$\text{dist}^{(2)}(b, e) = 2$$
$$\text{dist}^{(2)}(a, e) = 2$$
$$\text{dist}^{(2)}(c, e) = 1$$

from which we calculate the second-order closeness centrality of node e as $\text{ClC}^{(2)}(c) = 2$.

Again, in this particular example the second-order closeness centrality corresponds both to the temporal and the first-order closeness centrality. This is different in the second-order network shown in Figure 3.2b, which corresponds to the temporal network G_2 shown in Figure 3.1b. Here, we find that the second-order node $c - e$ can only be reached via a single shortest path $(a - c), (c - e)$ as well as via an additional second-order "path" of

length zero from $e - c$ to itself. From this, we can calculate the second-order distances

$$\text{dist}^{(2)}(a, e) = 2$$
$$\text{dist}^{(2)}(c, e) = 1$$

and for the second-order closeness centrality of node e we thus obtain $\text{ClC}^{(2)}(c) = 1.5$, which coincides with the temporal closeness of node e in the underlying temporal network G_2.

Using the the second-order closeness centrality introduced above, let us now study the correlations between the temporal and the second-order closeness centralities of nodes in our six data sets. The results of this analysis are shown in the right column of Table 4.2. For five of the six data sets we observe significantly larger correlation coefficients than those reported for the first-order closeness centrality in Table 4.2. The largest increase of the Pearson correlation coefficient from 0.91 to 0.97 is achieved for the (FL) data set, while we observe no improvement of the (already large) Pearson correlation coefficient of 0.98 for (LT). We further observe significant increases in the Kendall-Tau rank correlation coefficients for all of the studied data sets, except for (LT) for which it remains the same. For the ranking of nodes in (EM), we find that a ranking based on second-order closeness centralities increases the Kendall-Tau rank correlation with the ground truth temporal centralities from 0.79 to 0.92, thus better representing the relative importance of nodes in the temporal network.

4.2.3 Temporal reach centrality

Concluding this section we finally study *reach centrality*, another notion of path-based centrality that captures the number of nodes that can be reached from a node via paths up to given maximum length s [19]. For static networks, such as a first-order aggregate network, we define the *first-order* reach centrality of a node v as

$$\text{RC}^{(1)}(v, s) := \sum_{w \in V} \Theta(s - \text{dist}^{(1)}(v, w)) \tag{4.11}$$

where $\Theta(\cdot)$ is the Heaviside function, $\text{dist}^{(1)}(v, u)$ is the length of a shortest path from node v to u in the static, first-order network, and s is a parameter specifying up to which length paths should be considered. Clearly, the reach centrality $\text{RC}^{(1)}(v, s = 1)$ of a node v is equal to its out-degree while $\text{RC}^{(1)}(v, s = \infty)$ is equal to the subset of nodes to which v is connected via directed paths of any length.

A *temporal reach centrality* can again easily be defined based on the notion of shortest time-respecting paths, as well as the temporal distance function $\mathrm{dist}^{\mathrm{temp}}(v, w)$ defined in Eq. 4.3. Here, for a given maximum time difference δ and a given value s, we are interested in how many different nodes can be reached via shortest time-respecting paths which have at most length s. In analogy to Eq. 4.11, we can thus define the *temporal reach centrality* $\mathrm{RC}^{\mathrm{temp}}(v)$ of a node v as:

$$\mathrm{RC}^{\mathrm{temp}}(v, s) := \sum_{w \in V} \Theta(s - \mathrm{dist}^{\mathrm{temp}}(v, w)). \tag{4.12}$$

We want to highlight that with this definition of reach centrality, we focus on the temporal-topological characteristics introduced by the ordering of links, which is why base our definition on the *shortest* rather than the *fastest* time-respecting paths.

It is finally easy to see that a *second-order reach centrality* can be defined in analogy to second-order closeness centrality. For this, all we have to do is to replace the distance function in Eq. 4.11 by our previously defined second-order distance function, thus obtaining the following definition:

$$\mathrm{RC}^{(2)}(v, s) := \sum_{w \in V} \Theta(s - \mathrm{dist}^{(2)}(v, w)) \,. \tag{4.13}$$

Using a value of $s = 2$, we again exemplify these definitions using our two illustrative examples. Let us first calculate the first-order reach centrality of node a based on the first-order aggregate network shown in Figure 3.1c. Here we find that there are paths of at most length $s = 2$ from node a to the three nodes c, d and e, from which we conclude $\mathrm{RC}^{(1)}(a, s = 2) = 3$. For the temporal reach centrality of node a in the temporal network G_1 shown in Figure 3.1a, we observe that there are time-respecting paths of at most length $s = 2$ from node a to the three nodes c, e and d. We hence conclude $\mathrm{RC}^{\mathrm{temp}}(a, s = 2) = 3$, finding that for G_1 the temporal reach centrality again corresponds to the first-order reach centrality. Again, this is not the case for the temporal network G_2 shown in Figure 3.1b. Here, node a is only connected to the nodes c and e via time-respecting paths of up to length two, which means that we have $\mathrm{RC}^{\mathrm{temp}}(a, s = 2) = 2$.

For the second-order reach centrality of node a in the temporal network G_1 let us now consider the second-order aggregate network shown in Figure 3.2a. Based on the shortest paths in the second-order network, we first find that the node $a - c$ is connected to two nodes $c - d$ and $c - e$ via shortest paths of length one. Furthermore, we find an additional shortest path of length zero which connects the second-order node $a - c$ to itself. Again,

using our second-order distance function $\text{dist}^{(2)}$ here we find the distances

$$\text{dist}^{(2)}(a, c) = 1$$
$$\text{dist}^{(2)}(a, e) = 2$$
$$\text{dist}^{(2)}(a, d) = 2$$

from which we conclude that three nodes c, e and d can be reached via paths of length at most two. From this we calculate the second-order reach centrality of node a in G_1 as $\text{RC}^{(2)}(a, s = 2) = 3$. Applying the same arguments to the example network G_2 and the corresponding second-order aggregate network shown in Figure 3.2b, for the same three nodes we find the following second-order distances:

$$\text{dist}^{(2)}(a, c) = 1$$
$$\text{dist}^{(2)}(a, e) = 2$$
$$\text{dist}^{(2)}(a, d) = \infty$$

We thus obtain a second-order reach centrality of $\text{RC}^{(2)}(a, s = 2) = 2$ which corresponds to the temporal reach centrality of node a in G_2.

In the following, we use the temporal reach centrality defined above as ground truth, while studying how well it can be approximated by first-order and second-order reach centralities calculated from the first- and second-order time-aggregated networks respectively. Different from the analyses for betweenness and closeness centralities, here we must additionally account for the fact that the reach centrality can be calculated for different values of the maximum path length s. This implies that the Pearson correlation coefficient ρ and the Kendall-Tau rank correlation coefficient τ must be calculated for each value of s individually. The results of this analysis are shown in Figure 4.2, which shows the obtained values for ρ and τ for the correlations between i) the temporal and the first-order reach centralities (black lines), and ii) the temporal and the second-order reach centralities (orange lines) for each of the six data sets introduced above. Thanks to our choice of the maximum time difference δ, for all of our data sets both the underlying first- and second-order networks are strongly connected. Assuming that D is the diameter of the corresponding aggregate network, for all $s \geq D$ we thus necessarily arrive at a situation where the reach centralities of all nodes are identical. For the results in Figure 4.2 this implies that for any $s > D$ the correlation values are undefined since the first- (or second-)order centralities of all nodes are the same. We thus only plot the correlation coefficients τ and ρ for $s < D$, in which case they are well-defined.

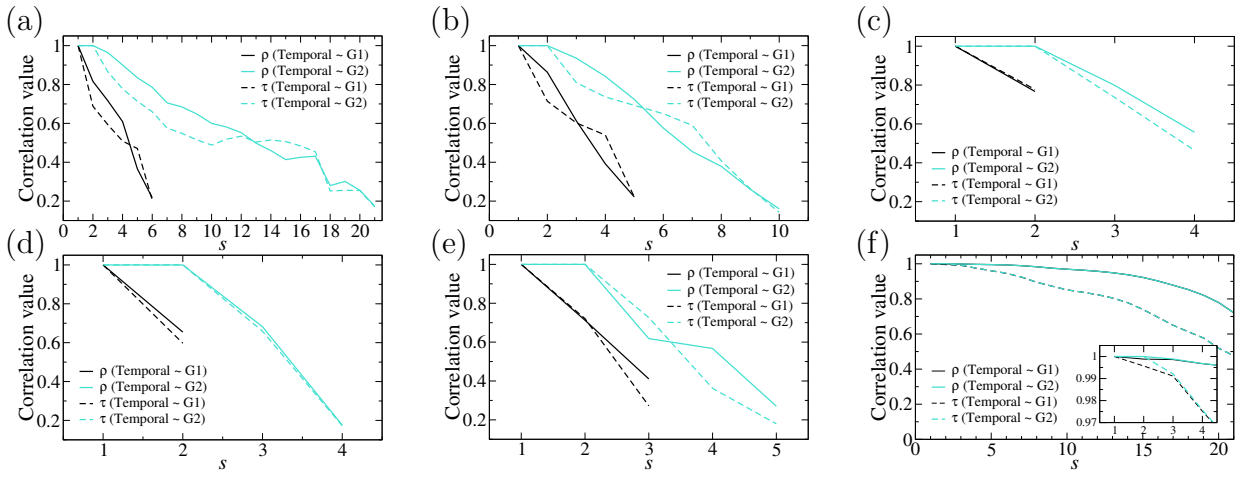


Figure 4.2: Correlation results for the reach centrality Pearson ρ and Kendall τ correlation coefficients between the temporal and the first-order reach centralities (black lines) and the temporal and the second-order reach centralities (orange lines) for (a) the Ants data set (AN), (b) the E-Mail data set (EM), (c) the Hospital data set (HO), (d) the Reality Mining dataset (RM), (e) the Flights data set (FL), and (f) the London Tube data set (LT). Inset: zoom to the area where there is a small deviation between values for the case of the London Tube data set.

For $s = 1$, the only time-respecting paths considered consist of single links, and thus the temporal reach centralities by definition exactly correspond to the reach centralities calculated from the first- and second-order topologies. Consequently, for $s = 1$ we have $\tau = 1$ and $\rho = 1$ both for the first- and the second-order reach centrality. For $s = 2$ there is, again by definition, no difference between the temporal and the second-order reach centralities however the correlation values for the first-order reach centrality decreases since the first-order aggregate network does not accurately represent the structure of time-respecting paths of length two. For values $s > 2$, ρ and τ decrease both for the first and the second-order centralities since neither representation can accurately represent time-respecting paths with lengths $s > 2$. However the results also highlight the important fact that second-order reach centralities better approximate temporal reach centralities for all values of $s > 2$.

We conclude this section by providing detailed results for the specific value of $s = 3$ shown in in Table 4.3. The choice of a parameter $s > 2$ means that for the second-order reach centrality we will not trivially obtain correlation values of 1 because we would only consider time-respecting paths of length two which are captured in the second-order aggregate network. However, since the diameter of the first-order aggregate network for two of our systems (RM and HO) is equal to three, we can only report results on the correlations between the temporal and the first-order reach centralities for four data sets.

	$RC^{temp} \sim RC^{(1)}$		$RC^{temp} \sim RC^{(2)}$	
	Pearson	Kendall-Tau	Pearson	Kendall-Tau
Ants (AN)	0.72 (8.23e-11)	0.59 (1.38e-11)	0.96 (9.50e-36)	0.86 (6.40e-23)
E-Mail (EM)	0.61 (3.17e-11)	0.60 (3.55e-18)	0.94 (2.74e-44)	0.81 (1.52e-31)
Hospital (HO)	NA	NA	0.80 (7.95e-13)	0.74 (7.23e-15)
RealityMining (RM)	NA	NA	0.68 (3.76e-09)	0.66 (2.78e-13)
Flights (FL)	0.41 (4.68e-06)	0.27 (1.46e-05)	0.62 (1.44e-13)	0.73 (6.53e-31)
London Tube (LT)	1.00 (4.65e-168)	1.00 (9.32e-64)	1.00 (1.92e-173)	1.00 (7.00e-64)

Table 4.3: Pearson and Kendall-Tau rank correlation coefficients between temporal reach centrality (ground truth) and reach centrality for $s = 3$ calculated based on the first–order aggregate network and the second-order aggregate network. Values in parentheses indicate the p-value.

Remarkably, for the (LT) data sets we observe a perfect correlation with the temporal reach centrality, which means that for this data set reach centralities are seemingly not affected by the temporal characteristics of the system. This is different for (FL), for which we observe a small Pearson correlation of $\rho = 0.41$, with an associated $\tau = 0.27$. These results show that, for the (FL) data set, temporal characteristics of the data do not allow temporal reach centralities to be approximated based on the first-order aggregate network. For the second-order reach centralities shown in the right columns of Table 4.3, we observe a significant increase in both the Pearson and the Kendall-Tau correlation coefficients for all of the data sets, except for (LT). The largest increase of the Pearson correlation coefficient is again obtained for (EM), increasing from 0.61 to 0.94 with an associated increase of the Kendall-Tau correlation coefficient from 0.60 to 0.81. We thus conclude that again, second-order reach centralities better capture the true (temporal) importance of nodes than a simple first-order approximation.

4.3 Conclusion

In summary, we have introduced a framework for the analysis of path-based notions of node centralities in temporal networks. In particular, we defined temporal versions of three path-based centrality measures which highlight the influence of the temporal-topological dimension introduced by the specific timing and ordering of time-stamped links in temporal networks. Using six data sets on real-world temporal networks, we have studied to what extent static notions of betweenness, closeness and reach centrality differ from their temporal counterparts. While for some data sets node centralities in the (first-order) time-

aggregated, static network can be used as reasonable proxies for temporal centralities, our results show that for other data sets this is not the case. Here we found that an analysis of time-aggregated static networks that neglect the time dimensions can yield misleading results about the importance of nodes.

In order to overcome these limitations, we have further used the framework of higher-order aggregate networks introduced in Chapter 3. The basic idea of this construction is that a k-th order aggregate network captures the statistics of time-respecting paths of length k, thus facilitating a higher-order analysis that incorporates both the topology and the ordering of links in temporal networks. We demonstrate the power of this framework through the definition of *second-order* centralities which can easily be calculated based on shortest paths in a second-order aggregate network. Despite the fact that these centralities can easily be calculated based on a simple static network structure, we find that the resulting second-order centrality measures capture better the centralities of nodes in the underlying temporal networks.

Closing, we would like to highlight a number of open issues which have to be considered in future works. First and foremost, all of our results have been obtained based on simple *unweighted* notions of centralities, even though in principle both the first-and second-order aggregate networks allow for the definition of link weights. Hence, our results have been obtained based on a rather simple perspective which does not incorporate the full information about path statistics preserved by our higher-order aggregate network abstraction. We thus expect a future extension to weighted higher-order aggregate networks to capture the true temporal centralities of nodes even more closely.

Moreover, while we can in principle define higher-order networks of any order k, in our work we have focused on second-order representations and the corresponding generalizations of path-based centralities. The choice to limit our study to $k = 2$ is mainly due to available data which, for the six temporal networks studied in this work, are not guaranteed to provide meaningful statistics for time-respecting paths with larger lengths k that are the basis for a k-th order aggregate network. Under what conditions higher-order aggregate networks with orders of $k > 2$ can help us to obtain even better approximations for temporal centralities is thus an open question that should be studied in the future.

Despite these open issues, we consider the fact that the simple second-order centrality measures introduced in our work already yield good approximations of the underlying temporal centralities which is a promising aspect of our framework. In this respect, second-order time-aggregated representations of temporal networks can be considered a simple, yet powerful abstraction for the higher-order analysis of time-stamped network data.

Finally, an important general question that arises in the analysis of time-stamped network data addresses the question under which conditions a time-aggregated analysis is sufficient, as opposed to a detailed analysis of time-stamped links and time-respecting paths. Thanks to their simplicity, computational efficiency, and the availability of software tools, time-aggregated analyses of *static node centralities* are popular and widely used throughout different disciplines. However, the results of our analysis, as well as of similar studies on dynamical processes, community structures and node centralities [88, 127, 140, 144] show that order correlations in real-world temporal networks crucially influence causality, thus potentially rendering such static analyses invalid. Through a calculation of *temporal node centralities* these temporal correlations can be included in the analysis of time-stamped data. However, especially for larger values of the maximum time difference δ, the extraction of all *shortest time-respecting* paths imposes computational costs that can be prohibitive for large data sets. A particular benefit of our approach is that the calculation of second-order centrality measures is computationally efficient as it merely requires i) the extraction of time-respecting paths of length two in the time-stamped data, and ii) the calculation of shortest paths in a *static* second-order network. Therefore, we argue that our approach is a simple and efficient (static) approximation of temporal centralities which, compared to a calculation of *first-order* centralities, nevertheless provides significant additional insights into the temporal dimension of complex systems.

Chapter 5

Temporal Causality: Slow-down or Speed-up

Summary

Recent research has highlighted limitations of studying complex systems with time-varying topologies from the perspective of static, time-aggregated networks. Non-Markovian characteristics resulting from the ordering of interactions in temporal networks were identified as one important mechanism that alters causality, and affects dynamical processes. So far, an analytical explanation for this phenomenon and for the significant variations observed across different systems is missing. Here we introduce a methodology that allows to analytically predict causality-driven changes of diffusion speed in non-Markovian temporal networks. Validating our predictions in six data sets, we show that - compared to the time-aggregated network - non-Markovian characteristics can lead to both a slow-down, or speed-up of diffusion which can even outweigh the decelerating effect of community and geodesic structures in the static topology. Thus, non-Markovian properties of temporal networks constitute an important additional dimension of complexity in time-varying complex systems.

Based on Scholtes, I., Wider, N., Pfitzner, R., Garas. A., Tessone, C.J. and Schweitzer, F. *Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks*, Nature Communications, vol. 5, number 5024, 2014. All authors conceived and designed the research and wrote the article. I.S. and N.W. analyzed the data, performed the simulations and provided the analytical results. Section 5.4.2 includes additional results by NW not part of the mentioned publication.

5.1 Introduction

The non-stationarity of the time-varying network topologies influences several aspects of complex systems. As shown in Section 4.2 properties that are directly related to the topology such as centrality can differ regarding a temporal or time-aggregated perspective. Several works have shown that compared to systems where, similar to static networks, most or all links are available concurrently, dynamical processes like epidemic spreading or diffusion are slowed down by the continuously switching topologies of temporal networks [100, 106, 133]. Other works show that the dynamics of network topologies can introduce noise which fosters certain types of consensus processes [8, 159].

Considering interactions in dynamic networks as a time series of events, a number of recent works focused on the question of whether observed inter-event times are consistent with the Poissonian distribution expected from a memoryless stochastic process. For a number of dynamic social systems, it has been shown that inter-event times follow non-Poissonian, heavy-tail distributions, and that the resulting bursty interaction patterns influence the speed of dynamical processes like spreading and diffusion [66, 70, 73, 77, 78, 125, 126, 134, 136, 151, 153]. More precisely, it was shown that such dynamical processes can be significantly slowed down with respect to their static representations. While all of these works highlight the importance of temporal information in the study of networks, there are a number of questions that have not been answered satisfactorily. Most empirical studies of dynamical processes in temporal networks focus on the influence of heavy-tail inter-event time distributions in dynamic social networks, which likely result from human task-execution mechanisms [49, 58, 72]. However, inter-event time distributions cannot explain temporality effects in other types of dynamic complex systems in which interactions are distributed homogeneously in time. Furthermore, this approach requires that sufficiently precise time stamps can be assigned to interactions, thus excluding path-based data where merely the ordering of interactions can be inferred. Recent works have shown that order correlations in temporal networks lead to *causality structures* which significantly deviate from what is expected based on paths in the corresponding time-aggregated networks [88, 127, 139].

Studying time-respecting paths $a \to b \to c$ from the perspective of a contact sequence a, b, c passing through node b, it was shown that the next contact c not only depends on the current contact b, but also on the previous one [86, 127, 139, 152, 156]. As a consequence, contact sequences in real-world temporal networks exhibit *non-Markovian characteristics* that are in conflict with the *Markovian* assumption implicitly made when studying temporal networks from the perspective of time-aggregated networks, and which

can neither be attributed to inter-event time distributions, nor to the concurrency or duration of interactions [86, 88, 127, 139]. Furthermore, it was shown that causality structures resulting from non-Markovian contact sequences influence both the speed of and the paths taken by dynamical processes [127, 139]. These works not only question the applicability of the static network paradigm when modeling dynamic complex systems, they also highlight a *temporal-topological* dimension of temporal networks which is ignored when exclusively focusing on time distributions of events and associated changes in the *duration* of dynamical processes. In line with the general lack of analytical approaches to understand and predict the effects of network dynamics on dynamical processes [100, 129], an analytical explanation for the influence of causality structures in real-world complex systems, as well as for the significant variations observed across different systems, is currently missing.

To fill these gaps, in this chapter we introduce an analytical approach that allows to study dynamical processes in non-Markovian temporal networks. In particular, we use higher-order time-aggregated representations of temporal networks to preserve causality, and use them to define Markov models for non-Markovian interaction sequences. We show that the eigenvalue spectrum of the associated transition matrices explains the slow-down and speed-up of diffusion processes in temporal networks compared to time-aggregated networks. We derive an analytical prediction for direction and magnitude of the change in a temporal network, validate it against six empirical data sets, and show that order correlations can both slow-down or speed-up diffusion even in systems with the same static topology. Our results highlight that non-Markovian characteristics of temporal networks can either enforce or mitigate the influence of topological properties on dynamical processes. As such, they constitute an important additional dimension of complexity that needs to be taken into account when studying time-varying network topologies.

5.2 Higher-order Markov models for temporal networks

In this section we focus on the causality nature of time-respecting paths. As it was pointed out in Section 3.3 higher-order aggregated networks are able to capture some of the causality inherent in temporal sequences. We especially focus on time-respecting paths of length two and refer to them as *two-paths*. Representing the shortest possible time-ordered interaction sequence, two-paths are the simplest possible extension of links (which can be viewed as "one-paths") that capture causality in temporal networks. As such two-paths are a particularly simple abstraction that allows to study causality in temporal networks [127, 140].

Our approach utilises a *state space expansion* to obtain a higher-order Markovian representation of non-Markovian temporal networks [109]. This means that a non-Markovian *sequence of interactions* in which the next interaction only depends on the previous one (i.e. one-step memory), can be modeled by a Markovian stochastic process that generates a *sequence of two-paths*.

5.2.1 Causality-preserving time-aggregated networks

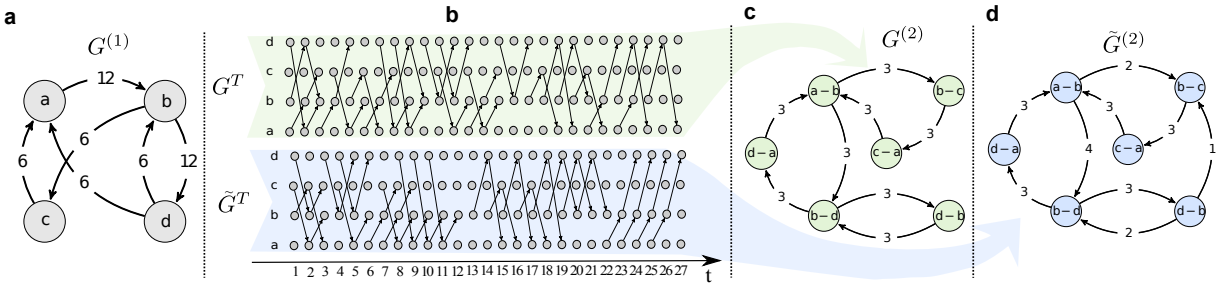


Figure 5.1: Two temporal networks with the same first-order, but different second-order time-aggregated networks (a) Time-aggregated network $G^{(1)}$, whose link weights capture the number of times each link occurred in a temporal network. The time-aggregated network is consistent with both temporal networks shown in (b). (b) Time-unfolded representations of two temporal networks, each consisting of four nodes and 27 time steps, both consistent with $G^{(1)}$. Differences in their causality structures are highlighted by the corresponding second-order aggregate networks shown in (c) and (d). Both second-order aggregate networks are consistent with $G^{(1)}$.

It is important to use a causality preserving model when dealing with temporal networks. Figure 5.1 (b) shows time-unfolded representations of two different temporal networks G^T and $\tilde{G}^T$ consisting of four nodes and 27 time steps. While both examples correspond to the same weighted time-aggregated network shown in Figure 5.1 (a), the two temporal networks differ in terms of the *ordering* of interactions. As a consequence, assuming a maximum time difference of $\delta = 1$, the time-unfolded representations reveal that a time-respecting path $d \to b \to c$ only exists in the temporal network $\tilde{G}^T$, while it is absent in G^T. This simple example illustrates how the mere ordering of interactions influences causality in temporal networks. In the following we present in detail how we use second-order time-aggregated networks as proxy for a causality-preserving approach.

Analogous to a *first-order time-aggregated network* $G^{(1)}$ consisting of (first-order) nodes $V^{(1)}$ and (first-order) links $E^{(1)}$, a *second-order time-aggregated network* $G^{(2)}$ consists of *second-order nodes* $V^{(2)}$ and *second-order links* $E^{(2)}$. Similar to a directed line graph construction [61], each second-order node represents a link in the first-order aggregate

network. As second-order link, we define all possible paths of length two in the first-order aggregate network, i.e. the set of all pairs (e_1, e_2) for links $e_1 = (a, b)$ and $e_2 = (b, c)$ in $G^{(1)}$. As outlined in Section 3.3 *second-order link weights* $w^{(2)}(e_1, e_2)$ can be defined as the relative frequency of time-respecting paths $(a, b; t_1) \to (b, c; t_2)$ of length two in a temporal network.

This approach allows us to generate a second-order network representation where each second-order node represents an link in the underlying temporal network, each second-order link represents a time-respecting path of length two, and weights $w^{(2)}$ capture the statistics of two-paths in the temporal network. An interesting aspect of this construction is that it allows to easily define *second-order Markov models* generating contact sequences which exhibit "one-step memory" and which thus correctly reproduce the statistics of time-respecting paths of length two in the original temporal network.

5.2.2 Transition matrices

In the following sections we will investigate a dynamical process based on a *random walk*. A random walker can start at any node of a network and moves to another node in every time step. The walker can usually only move to adjacent nodes in the direction of a link. The decision where the walker moves is random according to some given *transition probabilities* between the nodes. Recall Section 2.2.1 for more details. Given the relative frequencies of links in a temporal sequence one such transition probability can be defined in an empirical way. Considering a starting node b we can compute the empirical probability of any next link that is formed including b as source node. From a first-order perspective the probability for b to form a link with any other node c is equal to,

$$P(b \to c) = w^{(1)}(b, c) \left(\sum_{c' \in V^{(1)}} w^{(1)}(b, c') \right)^{-1} . \tag{5.1}$$

Hence, the transition probabilities from a first-order perspective is the same for all neighbors of b and 0 for all other nodes that were never connected to b in the temporal sequence. However, considering the causality stored in time-respecting paths the probability of a link $b \to c$ may dependent on a link $a \to b$ that occurred before $b \to c$. In a temporal network it does not necessarily hold that,

$$P(b \to c | a \to b) = P(b \to c | a' \to b) , \tag{5.2}$$

thus violating the *Markovian* assumption of first-order time-aggregated networks. In other words the probability of a link $b \to c$ is not necessarily the same for different preceding links, such as $a \to b$ and $a' \to b$. A second-order perspective takes into account previous links and therefore allows to define second-order transitions,

$$P(b \to c | a \to b) = w^{(2)}(a \to b, b \to c) \left(\sum_{c' \in V^{(2)}} w^{(2)}(a \to b, b \to c') \right)^{-1} . \tag{5.3}$$

A convenient way to store transition probabilities are so-called *transition matrices* denoted by $\mathbf{T}$. Similar to the adjacency matrix of a directed weighted network that stores the link weights for any link $\omega(a, b)$, the transition matrix stores the transition probability $P(a \to b)$ between nodes. In the following we apply this framework to second-order networks.

Using the second-order time-aggregated network $G^{(2)}$ and second-order link weights $w^{(2)}$ defined above, for all time-respecting paths $e_1 \to e_2$ of length two, with $e_1 = a \to b$ and $e_2 = b \to c$ we define the entries of the second-order transition matrix $\mathbf{T}^{(2)}$ for a random walk in the weighted network $G^{(2)}$ as

$$T^{(2)}_{e_1 e_2} := w^{(2)}(e_1, e_2) \left(\sum_{e' \in V^{(2)}} w^{(2)}(e_1, e') \right)^{-1} . \tag{5.4}$$

In line with the standard way of defining random walks on weighted networks [116], transition rates between nodes e_1 and e_2 are defined to be proportional to link weights and are normalised by the cumulative weight of all links (e_1, e') emanating from node e_1. If the transition matrix $\mathbf{T}^{(2)}$ is primitive, the Perron-Frobenius theorem guarantees that a unique leading eigenvector $\boldsymbol{\pi}$ of $\mathbf{T}^{(2)}$ exists. Note that $\mathbf{T}^{(2)}$ can always be made primitive by restricting it to the largest strongly connected component of $G^{(2)}$ and adding small positive diagonal entries.

While the transition matrix $\mathbf{T}^{(2)}$ captures the statistics of two-paths in a given temporal network, we can additionally define a maximum entropy transition matrix $\tilde{\mathbf{T}}^{(2)}$ which captures the two-path statistics one would expect based on the relative link weights in the first-order time-aggregated network. For $e_1 = (a, b)$ and $e_2 = (b, c)$, the entries $\tilde{T}^{(2)}_{e_1 e_2}$ corresponding to a two-path $e_1 \to e_2$ are given as

$$\tilde{T}^{(2)}_{e_1 e_2} := w^{(1)}(b, c) \left(\sum_{c' \in V^{(1)}} w^{(1)}(b, c') \right)^{-1} . \tag{5.5}$$

This second-order Markov model preserves the weights $w^{(1)}$ of links in $G^{(1)}$ and creates "Markovian" temporal networks in which consecutive links are independent from each other.

5.2.3 Entropy growth rate

The entropy of a second-order Markov model for a particular temporal network can be quantified in terms of the entropy growth rate of a transition matrix $\mathbf{T}^{(2)}$. This notion of entropy quantifies the amount of information that is lost about the current state of a Markov process based on a given transition matrix. We define the entropy growth rate of a second-order transition matrix as

$$H(\mathbf{T}^{(2)}) := -\sum_{e\in E^{(1)}} (\boldsymbol{\pi})_e \sum_{e'\in E^{(1)}} T^{(2)}_{ee'} \log_2 \left(T^{(2)}_{ee'}\right). \tag{5.6}$$

For a transition matrix which only consists of deterministic transitions with probability 1, the entropy growth rate is zero, while it reaches a (size-dependent) maximum for a transition matrix where every state can be reached with equal probability in every step.

From the perspective of statistical ensembles, which is commonly applied in the study of complex networks, each second-order transition matrix whose leading eigenvector $\boldsymbol{\pi}$ satisfies $(\boldsymbol{\pi})_e = w^{(1)}(a,b)$ ($\forall$ links $e = (a,b)$) defines a statistical ensemble of temporal networks constrained by a weighted time-aggregated network $G^{(1)}$ and a given two-path statistics. The entropy $H(\mathbf{T}^{(2)})$ of this ensemble can be defined as the entropy growth rate of the Markov chain described by the corresponding transition matrix [28]. Different from entropy measures previously applied to dynamic networks [180], this measure quantifies to what extent the next step in a contact sequence is determined by the previous one. For a specific second-order transition matrix $\mathbf{T}^{(2)}$ and a corresponding maximum entropy model $\tilde{\mathbf{T}}^{(2)}$, we define the *entropy growth rate ratio* as

$$\Lambda_H(\mathbf{T}^{(2)}) := H(\mathbf{T}^{(2)})/H(\tilde{\mathbf{T}}^{(2)}). \tag{5.7}$$

This ratio ranges between a minimum of zero for transition matrices corresponding to contact sequences that are completely deterministic, and a maximum of one for transition matrices corresponding to Markovian temporal networks. In general, an entropy growth rate ratio smaller than one highlights that the statistics of two-paths - and thus causality in the temporal network - deviates from what is expected based on the first-order aggregate network. As such, Λ_H is a simple measure that quantifies the importance of non-Markovian properties in temporal networks.

5.2.4 Example

To summarize this section, we illustrate our approach using the two temporal networks shown in Figure 5.1. Panels (c) and (d) show two second-order time-aggregated networks $G^{(2)}$ and $\tilde{G}^{(2)}$ corresponding to the temporal networks G^T and $\tilde{G}^T$ respectively. In particular, the absence of a time-respecting path $d \to b \to c$ in G^T is captured by the absence of the second-order link between the second-order nodes $e_1 = (d, b)$ and $e_2 = (b, c)$. Further differences between the causality structures of G^T and $\tilde{G}^T$ are captured by different second-order link weights. Notably, this example highlights that temporal networks giving rise to different second-order time-aggregated networks can still be consistent with the same first-order time-aggregated network.

In the following, we illustrate the construction of second-order transition matrices using the examples in Figure 5.1. For the second-order aggregate network $G^{(2)}$ shown in panel (c), corresponding to the temporal network G^T, the transition matrix $\mathbf{T}^{(2)}$ (rows/columns ordered as indicated) is

$$\mathbf{T}^{(2)} = \begin{array}{c|c} (a,b) & \\ (b,c) & \\ (b,d) & \\ (c,a) & \\ (d,a) & \\ (d,b) & \end{array} \left| \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \right. .$$

The leading eigenvector of a stochastic matrix captures the stationary distribution of the associated Markov chain. As such, the leading eigenvector $\boldsymbol{\pi}$ of the second-order transition matrix captures the stationary activation frequencies of links in contact sequences generated by the corresponding second-order Markov model. For the example above, we obtain a normalised leading eigenvector $\boldsymbol{\pi} = \left(\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$, which reproduces the relative weights of links in the first-order aggregate network shown in Figure 1 (a). In summary, interpreting $\mathbf{T}^{(2)}$ as transition matrix of a random walker in the second-order aggregate network, we obtain a second-order Markov model generating contact sequences that preserve the relative weights in the first-order aggregate network, as well as the statistics of two-paths. In line with recent observations that one-step memory is often sufficient to characterise time-respecting paths in empirical temporal networks [140], in the remainder of this chapter we focus on such second-order models. However, our findings can be generalised to n-th order networks $G^{(k)}$ and matrices $\mathbf{T}^{(k)}$ that capture the statistics of time-respecting paths of any length k. From this perspective, the weighted first-order aggregate network can be seen as a first-order approximation where weights only capture the

statistics of links, i.e. time-respecting paths of length one. Contact sequences generated by a random walk in the first-order time-aggregated network with transition probabilities proportional to link weights preserve the statistics of links but destroy the statistics of time-respecting paths. As such, a random walker in the first-order time-aggregate network must be interpreted as null model that destroys causality, and which can thus not be used to gain analytical insights about dynamical processes in non-Markovian temporal networks [10]. A second-order representation of the same null model can be constructed using a *maximum entropy second-order transition matrix* $\tilde{\mathbf{T}}^{(2)}$. For two links $e_1 = (a, b)$ and $e_2 = (b, c)$, the transition probability $\tilde{T}^{(2)}_{e_1 e_2}$ simply corresponds to the transition rate of a random walk across the weighted link (b, c) in the first-order aggregate network. This definition ensures that the corresponding random walker generates Markovian temporal networks which are consistent with a given weighted time-aggregated network, and which exhibit a two-path statistic as expected based on paths in the first-order aggregate network. We again illustrate our approach using the first-order time-aggregated network $G^{(1)}$ shown in the left panel of Figure 5.1. For this example, the transition matrix corresponding to a "Markovian" temporal network is given as

$$\tilde{\mathbf{T}}^{(2)} = \begin{array}{c|c} (a,b) & \left(\begin{array}{cccccc} 0 & 1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 & 0 & 0 \end{array} \right) \\ (b,c) & \\ (b,d) & \\ (c,a) & \\ (d,a) & \\ (d,b) & \end{array} .$$

Again, as leading eigenvector we obtain $\boldsymbol{\pi} = \left(\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$, confirming that the stationary activation frequencies of links correspond to the relative weights of links in the first-order time-aggregated network.

To quantify the importance of non-Markovian properties we apply the entropy growth rate ratio Λ_H to the example introduced in Figure 5.1. For the second-order transition matrices $\mathbf{T}^{(2)}$ and $\tilde{\mathbf{T}}^{(2)}$ we obtain $\Lambda_H(\mathbf{T}^{(2)}) = 0.84$ and thus $\Lambda_H(\mathbf{T}^{(2)}) < 1$. This confirms that $\mathbf{T}^{(2)}$ corresponds to a non-Markovian temporal network, and that the statistics of time-respecting paths in G^T deviates from what one could expect based on link frequencies in the first-order aggregate network. Considering the temporal network $\tilde{G}^T$, one easily verifies that link weights in the corresponding second-order aggregate network $\tilde{G}^{(2)}$ coincide with the transition matrix $\tilde{\mathbf{T}}^{(2)}$. The resulting entropy growth rate ratio of 1 for $\tilde{G}^T$ verifies that this temporal network does not exhibit non-Markovian characteristics and that two-path statistics do not deviate from what is expected based on the first-order aggregate network.

5.3 Causality-driven changes of diffusive behaviour

In the following, we highlight the relevance of causality in real-world systems by studying diffusion dynamics. We illustrate our results on the six empirical temporal network data sets described in Section 3.4. For each system, we study causality-driven changes of diffusion speed. In particular, we utilise a random walk process and study the time needed until node visitation probabilities converge to a stationary state [16, 93]. (See Section 2.2.1 for more details on random walks.) This convergence behaviour of a random walk is a simple proxy that captures the influence of both the topology and dynamics of temporal networks on general diffusive processes [116]. For a given convergence threshold ε, we compute a slow-down factor $\mathcal{S}(\varepsilon)$ which captures the slow-down of diffusive behaviour between the weighted aggregated network and a temporal network model derived from the empirical contact sequence respectively. In order to exclude effects related to node activities and inter-event time distributions and to exclusively focus on effects of causality observed in the real data sets, this model only preserves the weighted aggregate network as well as the statistics of two-paths in the data.

5.3.1 Diffusion dynamics in empirical temporal networks

We study causality-driven changes of diffusive behaviour in the six temporal network data sets (AN), (RM), (FL), (EM), (HO), (FL) and (LT) described above. We use the convergence behaviour of a random walk process as a proxy that captures the influence of both the topology and dynamics of temporal networks on general diffusive processes. For this, we first consider a random walk process in the weighted, time-aggregated network and study the time needed until node visitation probabilities converge to a stationary state. Starting from a randomly chosen node, in each step of the random walk the next step is chosen with probabilities proportional to the weights of incident links. A standard approach to assess the convergence time of random walks is to study the evolution of the *total variation distance* between observed node visitation probabilities and the stationary distribution [138]. For a distribution $\boldsymbol{\pi}_k$ of visitation probabilities $(\boldsymbol{\pi}_k)_v$ of nodes v after k steps of a random walk and a stationary distribution $\boldsymbol{\pi}$, the total variation distance is defined as

$$\Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) := \frac{1}{2} \sum_v | \, (\boldsymbol{\pi})_v - (\boldsymbol{\pi}_k)_v \, |.$$

For a given threshold distance ε, we define the convergence time $t_{agg}(\varepsilon)$ as the minimum number of steps k after which $\Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) < \varepsilon$. The random walk itineraries produced by

this simple random walk model correctly reproduce link weights in the time-aggregated network and the use of random walk itineraries as a model for temporal networks has been proposed before [10]. However, random walk itineraries do not preserve statistics of longer time-respecting paths and thus alter causality. In order to derive a causality-driven slow-down factor, we thus contrast the convergence time $t_{\text{agg}}(\varepsilon)$ with the convergence time $t_{\text{temp}}(\varepsilon)$ of a second model that additionally preserves the statistics of time-respecting paths of length two in the real data sets (see Section 3.4 for details on how we define time-respecting paths in the different data sets). Again starting with a random node, this model randomly chooses two-paths according to their relative frequencies in the data set, thus corresponding to a walk process which is advanced by two steps at a time. The random itineraries generated by this model correctly reproduce link weights in the time-aggregated network, and - different from a random walk in the time-aggregated network - the statistics of time-respecting paths of length two. For a given threshold distance ε, we again define the convergence time $t_{\text{temp}}(\varepsilon)$ as the minimum number of steps k after which $\Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) < \varepsilon$. For a convergence threshold ε, this allows us to define a causality-driven slow-down factor $\mathcal{S}(\varepsilon) := t_{\text{temp}}(\varepsilon)/t_{\text{agg}}(\varepsilon)$ that is due to the *temporal-topological* characteristics of time-respecting paths, while ruling out effects of inter-event time distributions or bursty node activities.

Figure 5.2 shows the causality-driven slow-down factor for the six empirical networks and different convergence thresholds ε. Even though the networks are of comparable size, deviations from the corresponding aggregate networks in the limit of small ε (i.e. the long-term behaviour) are markedly different. The values for the empirical slow-down factor are listed in table 5.1. The first four data sets signify a *slow-down* of diffusion, but for (FL) and (LT) we obtain a *speed-up* of diffusion by a factor of 1.04 and 4 respectively. It is not surprising that the travel patterns in (FL) and (LT) are "optimised" in such a way that diffusion is more efficient than in temporal networks generated by contacts between humans (RM, EM and HO) or ants (AN). However, an analytical explanation for the direction and magnitude of this phenomenon, as well as for the variations across systems, is a unsolved problem that we address in this chapter.

5.3.2 Predicting causality-driven changes of diffusion speed

A particularly interesting aspect of the second-order network representation introduced above is that *temporal transitivity* is preserved, i.e. the existence of two second-order links (e_1, e_2) and (e_2, e_3) implies that a time-respecting path $e_1 \to e_2 \to e_3$ exists in the underlying temporal network. Notably, the same is not true for first-order aggregate networks, which do not necessarily preserve temporal transitivity in terms of time-respecting

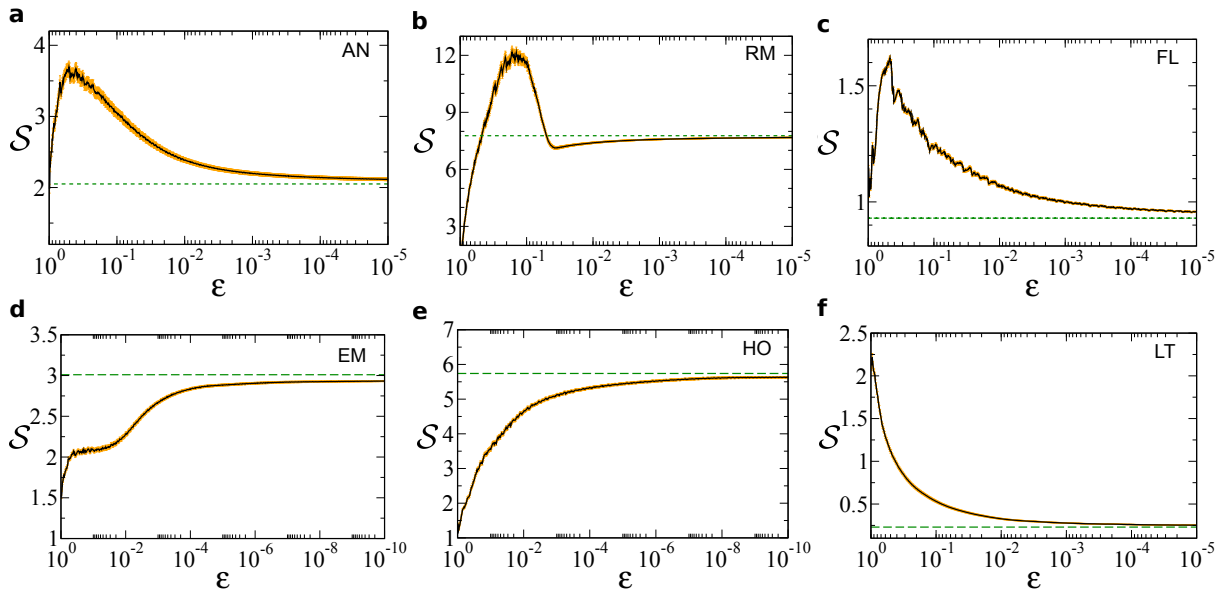


Figure 5.2: Causality-driven changes of diffusion speed We investigate the causality-driven changes of diffusive behaviour by an empirical study of the convergence time of random walks. For a given convergence threshold ε, we compute a slow-down factor $\mathcal{S}(\varepsilon)$ which captures the slow-down of diffusion in a temporal network model that preserves the weighted aggregate network as well as the statistics of time-respecting paths of length two in the data, compared to diffusion in the weighted aggregate network. The six panels show the ε-dependent slow-down factor for (a) the (AN) data set covering interactions between 89 ants, (b) the (RM) data on proximity relations between 64 academic staff members and students, (c) the (FL) data on flight itineraries connecting 116 airports, (d) the (EM) data covering E-Mail exchanges between 167 employees in a company, (e) the (HO) data on contacts between 75 patients and health-care workers in a hospital, and (f) the (LT) data on passenger journeys between 309 London Tube stations. Each result is the mean of random walks starting at every node, error bars indicate the s.e.m. (standard error of the mean). The predicted $\mathcal{S}^*$ value (see Eq. 5.8) is shown by the horizontal dashed line.

paths; i.e. the existence of two first-order links (a, b) and (b, c) does not imply that a time-respecting path $a \to b \to c$ exists. Transitivity of paths is a precondition for the use of algebraic methods in the study of dynamical processes. As such, it is possible to study diffusion dynamics in temporal networks based on the spectral properties of the matrix $\mathbf{T}^{(2)}$, while the same is not true for a transition matrix defined based on link weights in the first-order aggregate network. In particular, the convergence time of a random walk process (and thus diffusion speed) can be related to the second largest eigenvalue of its transition matrix [27]. For a primitive stochastic matrix with (not necessarily real) eigenvalues $1 = \lambda_1 > |\lambda_2| > |\lambda_3| \geq \ldots \geq |\lambda_n|$, one can show that the number of steps k after which the total variation distance $\Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi})$ between the visitation probabilities $\boldsymbol{\pi}_k$ and

Data Sets	$\mathcal{S}(\varepsilon)$	change	$\Lambda_H(\mathbf{T}^{(2)})$	$\mathcal{S}^*(\mathbf{T}^{(2)})$
Ants (AN)	$\approx 2.11 \pm 0.02, \quad \varepsilon = 10^{-5}$	slow-down	≈ 0.42	≈ 2.05
E-Mail (EM)	$\approx 2.93 \pm 0.005, \quad \varepsilon = 10^{-10}$	slow-down	≈ 0.62	≈ 3.01
Hospital (HO)	$\approx 5.63 \pm 0.019, \quad \varepsilon = 10^{-10}$	slow-down	≈ 0.71	≈ 5.75
RealityMining (RM)	$\approx 7.68 \pm 0.01, \quad \varepsilon = 10^{-5}$	slow-down	≈ 0.40	≈ 7.77
Flights (FL)	$\approx 0.957 \pm 0.002, \varepsilon = 10^{-5}$	speed-up	≈ 0.82	≈ 0.93
London Tube (LT)	$\approx 0.25 \pm 0.001, \quad \varepsilon = 10^{-5}$	speed-up	≈ 0.30	≈ 0.23

Table 5.1: Entropy growth rate ration Λ_H, slow down factor $\mathcal{S}(\varepsilon)$ and the analytical prediction of the slow-down factor $\mathcal{S}^*$ for the six empirical data sets that were investigated.

the stationary distribution $\boldsymbol{\pi}$ of a random walk falls below ε is proportional to $1/\ln(|\lambda_2|)$ (see Supplementary Note 1 for a detailed derivation). For a matrix $\mathbf{T}^{(2)}$ capturing the statistics of two-paths in an empirical temporal network, and a matrix $\tilde{\mathbf{T}}^{(2)}$ corresponding to the "Markovian" null model derived from the first-order aggregate network, an analytical prediction $\mathcal{S}^*$ for causality-driven changes of convergence speed can thus be derived as (see Section A.1 for more details)

$$\mathcal{S}^*(\mathbf{T}^{(2)}) := \ln(|\tilde{\lambda}_2|)/\ln(|\lambda_2|), \tag{5.8}$$

where λ_2 and $\tilde{\lambda}_2$ denote the second largest eigenvalue of $\mathbf{T}^{(2)}$ and $\tilde{\mathbf{T}}^{(2)}$ respectively. Depending on the eigenvalues λ_2 and $\tilde{\lambda}_2$, both a slow-down ($\mathcal{S}^*(\mathbf{T}^{(2)}) > 1$) or speed-up ($\mathcal{S}^*(\mathbf{T}^{(2)}) < 1$) of diffusion can occur.

This approach allows us to analytically study the effect of non-Markovian characteristics in the empirical data sets introduced in Section 3.4. For each data set we construct matrices $\mathbf{T}^{(2)}$ and $\tilde{\mathbf{T}}^{(2)}$ (see Eqs. 5.4 and 5.5), and compute the entropy growth rate ratio Λ_H for the corresponding statistical ensembles. The ratio indicates that the topologies of time-respecting paths in all six cases differ from what is expected from the first-order time-aggregated networks. The impact of these differences on diffusion can be quantified by means of the analytical prediction $\mathcal{S}^*(\mathbf{T}^{(2)})$. The obtained results can be found in Table 5.1.

All six predictions are consistent with the diffusion behaviour observed in numerical simulations in the limit of small ε (see Figure 5.2). As argued above, the significantly smaller magnitude of the slow-down effect in (AN) compared to (RM) can neither be attributed to differences in system size nor inter-event time distributions. A spectral analysis of $\mathbf{T}^{(2)}$ can explain the smaller slow-down of (AN) compared to (RM) by a "better connected" causal topology indicated by a smaller $\mathcal{S}^*$. Similarly, the large slow-down observed in (HO) can be related to a "badly connected" causal topology indicated by a large value of $\mathcal{S}^*$. For (FL), the analytical prediction $\mathcal{S}^*(\mathbf{T}^{(2)}) \approx 0.93$ is consistent with the asymp-

totic empirical speed-up observed in Figure 5.2. Similarly, the prediction $\mathcal{S}^*(\mathbf{T}^{(2)}) \approx 0.23$ for (LT) is in line with the speed-up observed in Figure 5.2. Here, the small value of $\mathcal{S}^*(\mathbf{T}^{(2)})$ highlights that the empirical second-order aggregate network is much better connected that one would expect from a Markovian temporal network, thus explaining the large speed-up by a factor of four. The non-linear behaviour of $\mathcal{S}(\varepsilon)$ can be understood by recalling that Eq. 5.8 makes the simplifying assumption that only λ_2 contributes to the convergence time, which holds in the limit of small ε. As ε increases, an increasing number of eigenvalues and eigenvectors have non-negligible contributions to the empirical slow-down $\mathcal{S}$.

5.4 Causality structures: slow-down or speed-up

Above, we showed that non-Markovian characteristics alter the causal topology of time-varying complex systems, and that the dynamics of diffusion in such systems can be explained by the resulting changes in the eigenvalue spectrum of higher-order aggregate networks, compared to the first-order aggregate network. We further analytically found that, depending on the system under study, non-Markovian characteristics can both slow-down or speed-up diffusion dynamics. In the following, we further investigate the mechanism behind the speed-up and slow-down by a model in which order-correlations can mitigate or enforce topology-driven limitations of diffusion speed.

5.4.1 Community structures

The model generates non-Markovian temporal networks consistent with a uniformly weighted aggregate network with two interconnected communities, each consisting of a random 4-regular graph with 50 nodes. A parameter $\sigma \in (-1, 1)$ controls whether time-respecting paths between nodes in *different* communities are - compared to a "Markovian" realisation - over- ($\sigma > 0$) or under-represented ($\sigma < 0$). The Markovian case coincides with $\sigma = 0$. An important aspect of this model is that realisations generated for any parameter σ are consistent with the same weighted aggregate network. The parameter σ exclusively influences the temporal ordering of interactions, but neither their frequency, topology nor their temporal distribution (see Section A.2 for model details and mathematical proofs). Figure 5.3 (a) shows the effect of σ on the entropy growth rate ratio Λ_H (blue, dashed line) and the predicted slow-down $\mathcal{S}^*$ (black, solid line). All non-Markovian realisations of the model (i.e. $\sigma \neq 0$) exhibit an entropy growth rate ratio $\Lambda_H < 1$ (blue dashed line) which signifies the presence of order correlations. Whether these correlations result in a

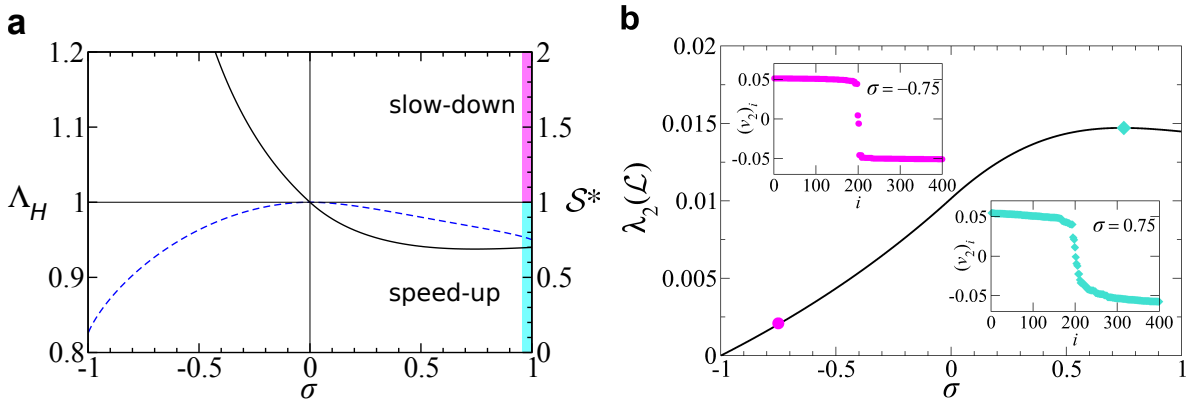


Figure 5.3: Slow-down and speed-up regimes in a temporal network model We analytically study a model of non-Markovian temporal networks consistent with a weighted first-order aggregate network with $n = 100$ nodes that has two pronounced communities with 50 nodes each. A parameter σ controls whether two-paths across communities are over-represented ($\sigma > 0$, turquoise) or under-represented ($\sigma < 0$, magenta) compared to the Markovian case ($\sigma = 0$). Realisations for different parameters only differ in the ordering of interactions and thus their second-order aggregate networks. Weighted first-order time-aggregated networks are the same for all parameters σ. (a) Entropy growth rate ratio Λ_H (blue, dashed line) and slow-down factor $\mathcal{S}^*$ (black, solid line) for different parameters σ. (b) Algebraic connectivity $\lambda_2(\mathcal{L})$ of the weighted second-order aggregate network for different parameters σ of non-Markovian temporal networks. Insets show the Fiedler vector for two points $\sigma = -0.75$ (magenta) and $\sigma = 0.75$ (turquoise) in the model's parameter space corresponding to cases where two-paths across communities are inhibited and enforced respectively.

speed-up ($\mathcal{S}^* < 1$) or slow-down ($\mathcal{S}^* > 1$) depends on how order correlations are aligned with community structures. For $\sigma < 0$, time-respecting paths across communities are *inhibited* and diffusion slows down compared to the time-aggregated network ($\mathcal{S}^* > 1$). For $\sigma > 0$, non-Markovian properties *enforce* time-respecting paths across communities and thus *mitigate the decelerating effect of community structures* on diffusion dynamics ($\mathcal{S}^* < 1$) [141]. We analytically substantiate this intuitive interpretation by means of a spectral analysis provided in Figure 5.3 (b). For each σ, we compute the algebraic connectivity of the causal topology, i.e. the second-smallest eigenvalue $\lambda_2(\mathcal{L})$ of the normalised Laplacian matrix $\mathcal{L} = \mathbf{I}_n - \mathbf{T}^{(2)}$ corresponding to the second-order aggregate network ($\mathbf{I}_n$ being the n-dimensional identity matrix). Larger values $\lambda_2(\mathcal{L})$ indicate "better-connected" topologies that do not exhibit *small cuts* [38, 177]. The effect of non-Markovian characteristics on $\lambda_2(\mathcal{L})$ validates that the speed-up and slow-down is due to the "connectivity" of the causal topology. In addition, the insets in Figure 5.3 (b) show entries $(\mathbf{v}_2)_i$ of the Fiedler vector, i.e. the eigenvector $\mathbf{v}_2(\mathcal{L})$ corresponding to eigenvalue $\lambda_2(\mathcal{L})$. The dis-

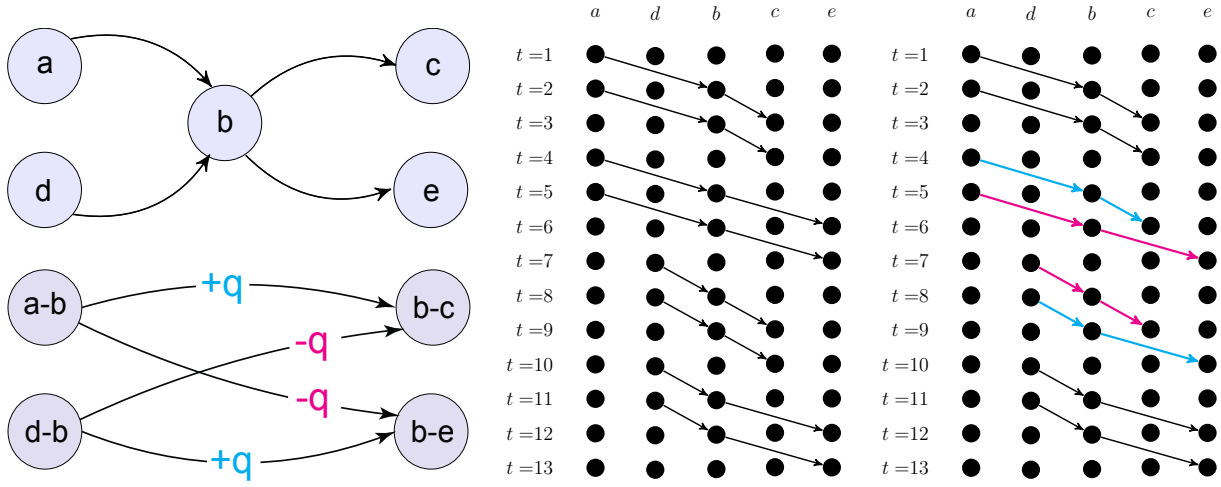
tribution of entries of $\mathbf{v}_2(\mathcal{L})$ is related to community structures and is frequently used for divisive spectral partitioning of networks [130]. For $\sigma = -0.75$, the strong community structure in the causal topology shows up as two separate value ranges with different signs, while the two entries close to zero represent links that interconnect communities. Apart from the larger algebraic connectivity, the distribution of entries in the Fiedler vector for $\sigma = 0.75$ shows that the separation between communities is less pronounced. This highlights that non-Markovian properties can effectively outweigh the decelerating effect of community structures in the time-aggregated network, and that the associated changes in the causality structures can be understood by an analysis of the spectrum of higher-order time-aggregated networks.

5.4.2 Geodesic structures

Another way to alter the diffusion speed in terms of "reordering" temporal links targets the geodesic structures of the network topology. The geodesic distance between two nodes a and b is equal to the length of the shortest path that connects them. In terms of a random walk it is the minimal amount of steps needed to reach b if the walker starts in a. The path *mitigating* and *enforcing* used to bridge communities in the model described in the previous section can in general be utilized to alter time-respecting paths in the desired way.

We start with a simple example of a 5 node network that is depicted in the top left in Figure 5.4. The node b has two incoming links from a and d and two outgoing links to c and e. The node b is the connecting node in this setup and each quantity coming from either a or d can travel to c or e. Let us consider that this network is the first-order aggregation of a temporal network. In this model we assume that that all links occur as often as needed but with the same frequency to assure that equal weights in the aggregation.

Let us consider that the initial temporal sequence is Markovian and therefore all paths have the same frequency. The second-order network contains four nodes $a-b$, $b-c$, $d-b$ and $b-e$ as depicted in the bottom left of Figure 5.4. The question arises how the links in the temporal sequence have to be reordered to change the frequency of second-orders link in a desired way. As an example let us assume that we want to increase the frequency of the two-path $a \to b \to c$ by some quantity q. This two-path consist of the links $a-b$ and $b-c$ that we have to acquire to build q new two-paths $a \to b \to c$. However, since the temporal sequence is assumed to be Markovian the only way to acquire additional links is to take them from another two-path. The only other two-path including $a-b$ is $a \to b \to e$ and the only other two-path including $b-c$ is $d \to b \to c$. Hence, if we



(a) First- and second-order networks

(b) Temporal sequence, original (left), reordered (right)

Figure 5.4: Concept of altering time-respecting path by reordering links Enforced paths (turquoise) get an increased frequency of q, while the frequency of other paths (magenta) have to be mitigated by the same amount. (a) First-order aggregate network on the top and second-order aggregate network on the bottom. (b) Time-unfolded representations of the original and reordered temporal network, both corresponding to the aggregated network shown in (a) top.

build q new two-paths $a \to b \to c$ the frequency of the aforementioned two-paths has to be decreased by q. On the other hand we also acquire q links $b - e$ and $d - b$ that are not anymore part of a two-path and can also be reused to build q additional two-paths $d \to b \to e$.

The reordering process is illustrated with a time-unfolded representation on the right in Figure 5.4. To summarized, through a reordering process that enforced the two-path $a \to b \to c$ by q we had to mitigate the two-paths $a \to b \to e$ and $d \to b \to c$ but also could enforce $d \to b \to e$ by the same quantity. The following link weights, corresponding to reordering links, preserve the first-order aggregate network and therefore the frequency of single links in the temporal sequence but alter the statistics of time-respecting paths,

$$\begin{aligned} w(a \to b \to c) + q \\ w(a \to b \to e) - q \\ w(d \to b \to c) - q \\ w(d \to b \to e) + q\,. \end{aligned} \tag{5.9}$$

Even though we used a simple example the concept itself can be applied to any kind of network. It just has to be assured that the total frequency of single links is preserved.

Therefore, enforcing a path also means that another path has to be mitigated to assure that the first-order time-aggregated remains unchanged.

A direct way to improve the diffusion speed in the network is to increase the flow between distant nodes. Recall that the algebraic connectivity $\lambda_2(\mathcal{L})$ is a proxy for the convergence rate of a random walk and therefore the diffusion speed inside a network. It was shown by Bojan Mohar [104] that,

$$\lambda_2(\mathcal{L}) \geq \frac{4}{n \cdot \mathrm{diam}(G)}\,, \tag{5.10}$$

where n is the amount of nodes and $\mathrm{diam}(G)$ is the diameter of the network G. The relation to network size and diameter imposes a lower bound on the algebraic connectivity. Lowering the diameter of the network increases the bound and therefore the potential for a fast convergence of the diffusion process. With temporal reordering of the network we are not able to alter the diameter of the first-order aggregated network. However, we still can alter the time-respecting paths and therefore the topology of the second-order networks. In specific cases we usually deal with directed links that are weighted. We can not apply Eqn. 5.10 directly to this framework but improving the flow between distant nodes will still increase the connectivity.

A simple way to improve the flow can be achieved by mitigating returning two-paths. By this we mean paths of the form $a \to b \to a$. In case of a random walk process such paths obviously slow down the spreading inside the network. Consider the simplest path $a \leftrightarrow b \leftrightarrow c$ that is connected in both direction and represents the basic component of any bidirectional path. We can apply equation 5.9 to enforce $a \to b \to c$ and $c \to b \to a$ and at the same time mitigating the returning two-paths $a \to b \to a$ and $c \to b \to c$. The returning two-paths can lead to significant slow-downs was also investigated in other studies [140].

Applying this procedure and therefore mitigating return steps of all two-paths in a bidirectional network can significantly speed-up the random walk process. We apply this idea to the six empirical data sets presented in Section 3.4. For each temporal sequence we first perform a second-order aggregation and detect all returning two-paths , i.e. paths of the form $a \to b \to a$. Considering equation 5.9 we randomly search for alternative two-paths such that the returning paths are mitigated and paths connecting three different nodes are enforced. The particular paths are chosen uniformly at random with no optimization criterion in mind. Therefore, there may occur situations where it is not possible to reorder returning two-paths in a desired way. In such cases we randomly choose another two-path that can be altered. We apply this reordering to all returning two-paths an perform the procedure multiple times on the same network and just report the best values that were

achieved. These are usually the cases where most returning paths could be reordered.

For each reordering we construct a second-order transition matrix $\hat{\mathbf{T}}^{(2)}$ and compute a slow-down factor in respect to the second-order aggregate network,

$$\mathcal{S}^{\#}(\hat{\mathbf{T}}^{(2)}) := \ln(|\lambda_2|)/\ln(|\hat{\lambda}_2|)\,. \tag{5.11}$$

The reordering is not compared to the "Markovian" null model, since we are interested in the change of the convergence rate compared to the actual two-path statistics. For (AN) we obtain $\mathcal{S}^{\#} = 0.49$, for (EM) we obtain $\mathcal{S}^{\#} = 0.62$, for (HO) we obtain $\mathcal{S}^{\#} = 0.37$ and for (RM) we obtain $\mathcal{S}^{\#} = 0.31$. Hence, for the 4 interaction data sets we observe significant speed-ups by mitigating returning paths compared to the actual path statistics. This implies that there are a lot of immediate repeated interaction between individuals in these data sets.

However, for the transportation data sets we obtain $\mathcal{S}^{\#} = 1$ for both (LT) and (FL). This is due to very few immediate returning steps in the networks. Both the London Tube and Flight data sets represent travel statistics where people navigate between distant places. They tend to travel across several nodes before they return. This effect is nicely capture in the London Tube flows illustrated in Figure 5.5. The figure shows the second-order aggregate network of the London Tube. Each node represent a link between two stations and each link a two-path. For each two-path we compute the frequency from a second-order aggregate perspective and a first-order aggregate perspective. The difference between the frequencies of the path is showed as a color scale between turquoise and magenta. Two-paths that are *more* frequent in the temporal sequence than assumed from an time-aggregate perspective are shown in turquoise. Two-paths that are *less* frequent in the temporal sequence than assumed from an time-aggregate perspective are shown in magenta. Interesting is the "rail" like structure in the network, highlighted by the inset of Figure 5.5. The turquoise links represent two-path connecting three different stations, while the magenta links represent returning two-paths. Hence, the nodes spanning the "rails" are reversed links as for example $a \to b$ and $b \to a$. The relation of such pairs propagate along the Tube lines and therefore highlight the efficiency of passenger flows in regard of traveling across distant parts of the network.

The method of enforcing and mitigating paths summarized in equation 5.9 is a general tool that can be applied in more elaborate ways. Here we only showed the impact on returning two-path therefore focusing on localized optimization of network flows. However, incorporating other parameters such as centrality measure, shortest paths or spanning trees should allow for even better techniques to alter the diffusion speed in temporal networks. Optimization methods should be applied in regard of particular applications

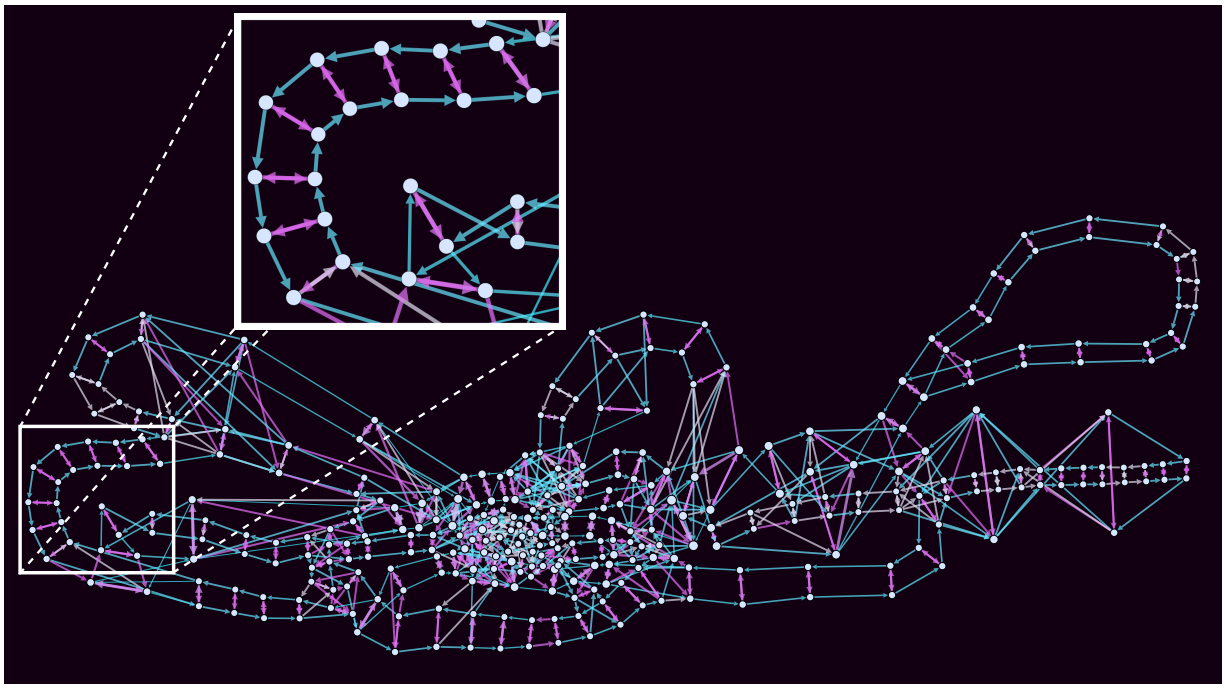


Figure 5.5: Second-order aggregated network of the London Tube (LT) Nodes represent links between London Tube stations, while links represent a two-paths between stations. Link colors indicate the difference in path statistics from the second-order aggregate network compared to a first-order aggregate network. Turquoise links are more frequent and magenta links less frequent in the temporal sequence than expect from an aggregate perspective. The inset shows, that the flows between distant nodes are more prominent then immediate returning itineraries.

to decide which properties are most significant for improving or hampering the diffusion in temporal system. Hence, the framework of higher-order aggregate networks is a useful tool that helps to design systems in such a way that the ordering of interactions and connections alters the diffusion speed in a desired way.

5.5 Conclusion

The abstraction of higher-order aggregate representations allows to define Markov models generating statistical ensembles of temporal networks that preserve the weighted aggregate network as well as the statistics of time-respecting paths. Focusing on second-order Markov models, we showed how transition matrices for such models can be computed based on empirical contact sequences. The ratio of entropy growth rates (see Eq. 5.7) between this transition matrix and that of a null model, which can easily be constructed from the

first-order aggregate network, allows to assess the importance of non-Markovian properties in a particular temporal network. Considering six different empirical data sets, we showed that spectral properties of the transition matrices capture the connectivity of the causal topology of real-world temporal networks. We demonstrate that this approach allows to analytically predict whether non-Markovian properties slow-down or speed-up diffusive processes as well as the magnitude of this change (see Eq. 5.8). With this, we provide the first *analytical explanation* for both the direction and magnitude of causality-driven changes in diffusive dynamics observed in empirical systems. Focusing on the finding that non-Markovian characteristics of temporal networks can both slow-down or speed-up diffusion processes, we finally introduce a simple model that allows to analytically investigate the underlying mechanisms. Our results show that the mere ordering of interactions can either mitigate or enforce topological properties that limit diffusion speed. Both our empirical and analytical studies confirm that causality structures in real-world systems have large and significant effects, slowing down diffusion by a factor of more than seven in one system, while other systems experience a speed-up by a factor of four compared to what is expected from the first-order time-aggregated network. These findings highlight that the causal topologies of time-varying complex systems constitute an important additional temporal dimension of complexity, which can reinforce, mitigate and even outweigh effects that are due to topological features like, e.g., community and geodesic structures.

Different from studies exclusively considering how interactions are distributed in time, in our study we focus on how their ordering influences causality structures in temporal networks. The finding that causality structures alone can lead to both a speed-up or a slow-down of diffusion highlights that, in order to understand the influence of the temporal dynamics in real-world systems, effects of *both* activity patterns and causality must be taken into account. Considering temporal networks in which interactions are homogeneously distributed in time, our approach further provides a novel explanation for changes in dynamical processes that cannot be explained in terms of bursty node activity patterns. An additional benefit of our approach is that it can be used for the network-based study of systems for which causal relations between different links can be inferred even though links cannot be assigned absolute time stamps. The data on airline and subway passenger itineraries that we analysed are two examples for such systems where only the ordering of links is known.

Our approach of constructing higher-order Markov models that preserve the statistics of time-respecting paths allows to study the *temporal-topological* dimension of time-varying complex systems - a dimension that is often ignored when exclusively focusing on changes in the *duration* of dynamical processes. The higher-order time-aggregated networks introduced in Chapter 3 are simple static representations of temporal networks which -

compared to first-order aggregate networks - better preserve causality. This approach provides interesting perspectives not only for analytical studies of further classes of dynamical processes in complex systems with time-varying interaction topologies. It is also a promising approach for the development of novel *temporal community detection algorithms* using, e.g., spectral clustering or random walk based methods as well as for the design of refined eigenvector-based centrality measures taking into account the ordering of links in dynamic networks. Finally, we foresee applications in the development of novel temporal network visualization methods, such as layout algorithms that make use of both the first- and the second-order time-aggregated networks.

Part II

Interconnectivity

"The only simple truth is that there is nothing simple in this complex universe. Everything relates. Everything connects."

Johnny Rich
The Human Script

Chapter 6

Lack of Information in Multi-layer Networks

Summary

We study properties of multi-layered, interconnected networks from an ensemble perspective, i.e. we analyze ensembles of multi-layer networks that share similar aggregate characteristics. Using a diffusive process that evolves on a multi-layer network, we analyze how the speed of diffusion depends on the aggregate characteristics of both intra- and inter-layer connectivity. Through a block-matrix model representing the distinct layers, we construct transition matrices of random walkers on multi-layer networks, and estimate expected properties of multi-layer networks using a mean-field approach. In addition, we quantify and explore conditions on the link topology that allow to estimate the ensemble average by only considering aggregate statistics of the layers. Our approach can be used when only partial information is available, like it is usually the case for real-world multi-layer complex systems.

Based on Wider, N., Garas, A., Scholtes, I. and Schweitzer, F. *Ensemble perspective on multi-layer networks* a chapter in *Interconnected Networks*, pp 37-59, Springer, 2016. NW conceived the study and wrote the article together with the other authors. Additionally, NW performed the simulations and provided the analytical results.

6.1 Introduction

In a multi-layered network each individual layer contains a network that is different from the networks contained in other layers, and the layer interconnectivity refers to the fact that nodes in different layers can be connected to each other. Nevertheless, it is often possible to extend and apply methods developed for single-layer (isolated) networks to multi-layer networks, assuming that all layers and the connections between them are known precisely. Unfortunately, when creating networks using relational data on real-world systems we are often confronted with situations where we *lack information* about the details of their multi-layer structure. In such situations, ensemble-based approaches allow us to reason about the expected properties of such networks, provided that we have access to aggregate statistics which can be used to define a statistical ensemble.

For instance, there are situations in which we are able to precisely map the topology *within* each layer individually, but we may not be able to obtain the detailed topology of connections *across* different layers. As an example, we may consider the topology of connections between users in different online social networks (OSNs). Such a system can be represented as a multi-layer network, where each layer represents the network of connections between users *within* one OSN. In addition, cross-layer connections are due to users which are members of multiple OSNs at the same time, and which can thus drive the dissemination of information across OSNs. Data on the network topology within particular OSNs are often readily available, however it is in general very difficult to identify accounts of the same user in different OSNs.

Contrary to the situation described above, we may also consider situations in which detailed information on the topology of cross-layer links is available, while the detailed topology of connections *within* layers is not known. For example, there may be a rather small number of static links *across* layers, while the topology of links *within* layers is too large and too dynamic to allow for a detailed mapping. Again, in such a situation we may still have access to partial, aggregate information on the *inter-layer* connectivity (such as the number of nodes or the density of links) which we can use in order to reason about a multi-layer complex system.

Both the above situations lead to multi-layer networks, and both require us to reason about a system with incomplete information. This problem can be addressed from a macroscopic perspective using *statistical ensembles*, and in this work we extend the ensemble perspective to multi-layer networks, where we have access to mere aggregate statistics either on links *within* or *across* layers. Combining both detailed and aggregate information on the links in a multi-layer network, we first define a statistical ensemble, i.e. a probability

space containing all network realizations that are consistent with available information. Secondly, we assume a probability mass function which assigns a probability to each possible realization in the ensemble. And finally, using either analytical or numerical techniques, we use the resulting probability space to reason about the *expected properties* of a network given that it is drawn from the ensemble.

The rest of the chapter is structured as follows. In Section 6.2 we present our methodological approach to model ensembles of multi-layer networks, we formally introduce the diffusion process that is assumed to run on the multi-layer network, and we introduce a method that allows to aggregate the statistics of links inside layers and across layers. In Section 6.3 we introduce a mean-field approach to approximate ensemble averages, and we investigate under which conditions it can be used to argue about diffusion in multi-layer networks. In particular we discuss three distinct cases according to different levels of information that we may have about the topology of links across the layers or inside the layers.

6.2 Methods and definitions

In our analysis we investigate a diffusion process that evolves on a static multi-layer network. More precisely we focus on diffusion dynamics modeled by a random walk process. Recall the definitions of a discrete random walk process introduced in Section 2.2.1, that we briefly summarize in the following.

A random walker can start at an arbitrary node in the network and in each time step moves to an adjacent node. Given a network $\mathbf{G}$ we can define a transition matrix $\mathbf{T}$ that contains the transition probabilities. The transition probability indicates with which likelihood a certain link is chosen. The position of the walker can be tracked in each step which results in a probability distribution π^t to find the walker at a particular node after t time steps. If the network is $\mathbf{G}$ strongly-connected and aperiodic, for $t \to \infty$ the probability distribution π^t converges to a stationary state π^*. The amount of time steps needed until π^t converges to π^* given some threshold ε can be estimated by second-largest eigenvalue λ_2 of $\mathbf{T}$. Therefore in the following we use $\lambda_2(\mathbf{T})$ as a proxy to measure and quantify the convergence behavior of a random walker on a network. An eigenvalue λ_2 close to one implies slow convergence, while λ_2 close to zero implies fast convergence.

In this chapter we will investigate a random walk process on a multi-layer network and analyze its convergence speed depending on the multi-layer topology.

6.2.1 Multi-layer network

The purpose of our study is to investigate diffusion processes on ensembles of networks with multiple interconnected layers. Thus, in the following we briefly recall the notion of multi-layer networks used in this chapter and introduced in Section 2.4. Let us consider a multi-layer network denoted by $\mathbf{G}$ that consist of L non-overlapping layers $G_1, \ldots, G_L$. Each of these layers G_l is a single-layer network $G_l = (V_l, E_l)$ where $V(G_l)$ and $E(G_l)$ denote the nodes and links of layer l respectively. We call the links $E(G_l)$ between nodes *within* the layers l *intra*-links. The multi-layer network $\mathbf{G}$ consist in total of n nodes, where $n = \sum_{l=1}^{L} |V(G_l)|$. In addition, we assume a set $E_I(\mathbf{G})$ of *inter-layer* links which connect nodes *across* layers, i.e. for each link $(u, v) \in E_I$ we have $u \in V(G_i)$ and $v \in V(G_j)$ for $i \neq j$. Inter-layer links induce a multipartite network with the independent sets $G_1, \ldots, G_L$.

In our study we consider undirected and unweighted networks, however some of our results may hold even for directed or weighted networks. Furthermore, from the perspective of a random walk process, we assume that inter- and intra-layer links are indistinguishable, i.e. transitions are made purely randomly irrespective of the type of link. As such, the multi-layer network can also be viewed as a huge single network consisting of subnetworks $G_1, \ldots, G_L$.

As mentioned above diffusion dynamics on networks can be studied analytically using transition matrices of random walkers [33, 93]. The multi-layer structure of a network can explicitly be incorporated in a random walk model by constructing a so-called *supra*-transition matrix [55, 148] similar to the supra-adjacency matrix used in [32, 33, 82]. The supra-adjacency matrix of a multi-layer network $\mathbf{G}$ can be defined in a block-matrix form as

$$\mathbf{A} = \left(\begin{array}{c|c|c|c|c} \mathbf{A}_1 & \ldots & \mathbf{A}_{1t} & \ldots & \mathbf{A}_{1L} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \mathbf{A}_{s1} & \ldots & \mathbf{A}_{st} & \ldots & \mathbf{A}_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \mathbf{A}_{L1} & \ldots & \mathbf{A}_{Lt} & \ldots & \mathbf{A}_{L} \end{array}\right). \tag{6.1}$$

On the diagonal we have the adjacency matrices $\mathbf{A}_1, \ldots, \mathbf{A}_L$ corresponding to the layers $G_1, \ldots, G_L$, thus entries of these block matrices represent the *intra-layer links* of the multi-layer network. Off-diagonal matrices $\mathbf{A}_{ij}$ for $i, j \in \{1, \ldots, L\}$ with $i \neq j$ represent *inter-layer* links that connect nodes in layer G_i to nodes in layer G_j. Since we consider undirected networks we have $\mathbf{A}_{ij}^{\top} = \mathbf{A}_{ji}$.

Based on a supra-adjacency matrix $\mathbf{A}$ we can easily define a *supra-transition* matrix $\mathbf{T}$ of

a random walker on a multi-layer network $\mathbf{G}$. In block-matrix form such a matrix can be written as:

$$\mathbf{T} = \left(\begin{array}{c|c|c|c|c} \mathbf{T}_1 & \dots & \mathbf{T}_{1t} & \dots & \mathbf{T}_{1L} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \mathbf{T}_{s1} & \dots & \mathbf{T}_{st} & \dots & \mathbf{T}_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \mathbf{T}_{L1} & \dots & \mathbf{T}_{Lt} & \dots & \mathbf{T}_{L} \end{array} \right) . \tag{6.2}$$

Here, each entry T_{ij} is defined as:

$$T_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{ik}} , \tag{6.3}$$

where a_{ij} are the corresponding entries of the supra-adjacency matrix $\mathbf{A}$. Note that, due the presence of inter-layer links, block matrices $\mathbf{T}_{ij}$ are in general not equal to the row-normalized version of block matrices $\mathbf{A}_{ij}$. The supra transition matrix defined above can be used to model a random walk process on a multi-layer network.

From an analytical perspective the supra-transition matrix can be treated in the same way as the transition matrix of a single layer as explained above. In the case of undirected networks the eigenvalues of a transition matrix are related to the eigenvalues of the normalized Laplacian matrix. In our case we study the second-largest eigenvalue of the supra-transition matrix and use it as a proxy for the efficiency of a network with respect to a diffusion process as pointed out above.

Using $\mathbf{T}$ we are able to model a diffusion process on a multi-layer network. Since we especially want to emphasize the relevance of the inter-links, in the next section we introduce a transition matrix that only considers transitions across layers and not between individual nodes. As we will see later, this aggregated transition matrix is useful to distinguish the influence of inter-layer and intra-layer links on the convergence behavior of a random walk process.

6.2.2 Multi-layer aggregation

The supra-transition matrix $\mathbf{T}$ introduced previously contains transition probabilities for any pair of nodes in the multi-layer network. In this sense $\mathbf{T}$ could also be the transition matrix of a large network, which is not divided in separate layers. In order to understand the effects of a layered structure, in this section we focus explicitly on transitions across layers. To do this we aggregate the statistics of inter-links and the intra-links of all single layers, i.e. we only consider the number of links inside an between layers. Thus, we

homogenize all individual nodes that belong to the same layer, since there no distinguishable from an aggregate perspective. This way we reduce the supra-transition matrix $\mathbf{T}$ of dimension n to an aggregated transition matrix $\mathfrak{T}$ of dimension L. We call this process *multi-layer aggregation* and the matrix $\mathfrak{T}$ the *layer-aggregated* or just *aggregated* transition matrix. Later on we will provide a relation between the eigenvalues of $\mathbf{T}$ and $\mathfrak{T}$, which will allow us to decompose the spectrum of $\mathbf{T}$. This is important since the convergence behavior of a random walk process depends on the second largest eigenvalues of $\mathbf{T}$.

Let us begin by discussing the construction process of the layer-aggregated transition matrix. Our goal is to define transition probabilities across any two layers G_s and G_t by averaging the transitions between any two nodes of G_s and G_t. Under certain conditions which will be specified in the following, these average transition probabilities can be representative for all nodes of the different layers.

Let $\mathbf{G}$ be a multi-layer network that consists of L layers $G_1, \ldots, G_L$. The transition probability to go from node v_i to any node v_j in $\mathbf{G}$ is defined as

$$P(v_i \to v_j) = \frac{\omega(v_i, v_j)}{\sum_k \omega(v_i, v_k)} \tag{6.4}$$

where $\omega(v_i, v_j)$ is the weight of a link connecting v_i with v_j. This is a general formalism, but since we only consider unweighted networks we have $\omega(v_i, v_j) = 1$ if and only if there is a link between the nodes v_i and v_j.

For each node v_i in layer G_s we require that the transition probabilities $P(v_i \to *)$ to nodes in another layer G_t fulfill the following equation

$$\alpha_{ss} \sum_{v_j \in V(G_s)} P(v_i \to v_j) = \alpha_{st} \sum_{v_k \in V(G_t)} P(v_i \to v_k) \quad \forall v_i \in V(G_s)\,, \tag{6.5}$$

where α_{st} is a factor that only depends on the layers G_s and G_t. The factor α_{ss} is used to normalize the transitions, such that $\sum_t \alpha_{st} = 1$ is satisfied. In other words Eq.(6.5) implies that the probability for a random walker at node i to stay inside layer G_s is a multiple of the probability to switch to layer G_t.

We can see that α_{st} isx independent of i, and therefore $\mathbf{T}_{st} = \alpha_{st}\mathbf{R}_{st}$ where $\mathbf{R}_{st}$ is a row stochastic matrix. This means that $\mathbf{T}_{st}$ resembles a scaled transition matrix, and α_{st} represents the weighted fraction of all links starting in G_s that end up in G_t. Thus, we

can define the aggregation of a supra-adjacency matrix satisfying Eq.(6.5) as

$$\mathfrak{T} = \left(\begin{array}{c|c|c|c|c} \alpha_{11} & \dots & \alpha_{1t} & \dots & \alpha_{1L} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \alpha_{s1} & \dots & \alpha_{st} & \dots & \alpha_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \alpha_{L1} & \dots & \alpha_{Lt} & \dots & \alpha_{LL} \end{array}\right) . \tag{6.6}$$

If a multi-layer network $\mathbf{G}$ satisfies Eq.(6.5) we can follow that the spectrum of the aggregated matrix $\mathfrak{T} = \{\alpha_{st}\}_{st}$ is

$$Spec(\mathfrak{T}) = \{1, \lambda_2, \dots, \lambda_L\}\,, \tag{6.7}$$

and it holds that $\lambda_2, \dots, \lambda_L \in Spec(\mathbf{T})$ (see Prop. 1 in the Appendix B.1).

This relation implies that the aggregated matrix $\mathfrak{T}$ preserves L eigenvalues of the supra-transition matrix $\mathbf{T}$, where L is the amount of layers. In other words, under the condition that Eq.(6.5) holds, we are able to make statements about the spectrum of the transition matrix $\mathbf{T}$ only using the layer-aggregated transition matrix $\mathfrak{T}$.

Similar to the Fiedler vector, i.e. the eigenvector corresponding to the second smallest eigenvalue of the Laplacian matrix, here we may use the eigenvector v_2 corresponding to the second largest eigenvalue λ_2 of the transition matrix $\mathbf{T}$. The vector v_2 contains negative and positive entries and sums up to zero. If all individual nodes that belong to the same layer correspond to entries of v_2 with the same sign, we consider the layers of $\mathbf{G}$ partitioned according to v_2, which is also called spectral partitioning or spectral bisection [38, 39]. In this case, according to Cor. 1 in the Appendix B.1, it holds that $\lambda_2(T) = \lambda_2(\mathfrak{T})$.

We note that the multi-layer aggregation, performed according to a spectral partitioning, has similarities to spectral coarse-graining [52]. The multi-layer aggregation presented here decreases the state space as well, but still preserves parts of the spectrum.

The spectral properties introduced in this section are important for our ensemble estimations that follows, since we characterize the diffusion process by its convergence efficiency measured through the second-largest eigenvalue $\lambda_2(\mathbf{T})$ of the supra-transition matrix. However, as outlined before, if Eq.(6.5) holds then this eigenvalue is equal to the second-largest eigenvalue $\lambda_2(\mathfrak{T})$ of the aggregated transition matrix $\mathfrak{T}$. Considering that for the construction of $\mathfrak{T}$ we only used aggregated statistics on the network and not the detailed topologies of the inter-links or any of the intra-links of all single layers, this already provides a hint how we can treat a system in the case of limited information.

6.3 Mean-field approximation of ensemble properties

With the layer aggregation introduced in the previous section, we are now able to deal with multi-layer network ensembles in case of limited information. In our case, this information concerns knowledge either of the inter-link topology between layers or the intra-link topologies of all single layers. For our purpose we define ensembles based on the inter-link densities and intra-link densities of all single layers, more precisely, by using the amount of nodes, the amount of inter-links across any two layers, and the amount of intra-links of all single layers. The number of nodes in individual layers are represented by the vector $\vec{n} = \{n_1, \ldots, n_L\}$ and the number of links between layers by a matrix $\mathbf{M}$ with entries m_{st} where s gives the source layer and t the target layer. Intra-layer links have both of their ends in the same layer and therefore we assume that the diagonal elements m_{ss} are equal to the amount of desired intra-links multiplied by two. We denote the ensemble defined by these two quantities $\mathcal{E}(\vec{n}, \mathbf{M})$.

A single random realization of this ensemble satisfies the aggregated statistics given by $\mathbf{M}$ and $\vec{n}$. We assume a random uniform distribution of links and therefore each realization of $\mathcal{E}(\vec{n}, \mathbf{M})$ has the same probability. However, instead of single realizations we are rather interested in the average values of all possible realizations. For each multi-layer network realization $\mathbf{G}$ of $\mathcal{E}(\vec{n}, \mathbf{M})$ we build the supra-transition matrix $\mathbf{T}$, which defines a random walk process that is different for every realization. As discussed above, a proxy of the convergence quality of these random walk processes is given by the second-largest eigenvalue $\lambda_2(\mathbf{T})$. Our goal is to estimate the average λ_2 of the ensemble $\mathcal{E}(\vec{n}, \mathbf{M})$, and we do this using a mean-field approach on the supra-transition matrix $\mathbf{T}$ that is similar to Refs. [57, 99].

Hereafter we will provide a mean-field approach for the general case, i.e. when the exact topology of inter-links and intra-links of all single layers are unknown. Next, building on this approach, we will discuss the case for which we have full knowledge of the intra-link topology but we do not know the inter-link topology, and the case for which we have full knowledge of the inter-link topology but we do not know the intra-link topology.

6.3.1 Case I: unknown inter- and intra-connectivity

For this case we only assume knowledge of the ensemble parameters $\mathbf{M}$ and $\vec{n}$. We define a mean-field adjacency matrix $\hat{\mathbf{A}}$ with a block structure similar to Eq.(6.1), and for each $\hat{\mathbf{A}}_{st}$ we are only given the amount of links equal to m_{st}. Since we do not know how these links are assigned to the entries $\mathbf{A}_{st}$, without loss of generality we assume a uniform distribution.

Thus, for the blocks of $\hat{\mathbf{A}}$ we have

$$\hat{\mathbf{A}}_{st} = \left\{ \frac{m_{st}}{n_s n_t} \right\}_{ij}, \quad i \in \{1, \ldots, n_s\}, \quad j \in \{1, \ldots, n_t\}. \tag{6.8}$$

Following the discussion of Section 6.2, based on the mean-field adjacency matrix we define a mean-field transition matrix $\hat{\mathbf{T}}$. The transition probability between any two nodes $i, j \in G_s$ for a fixed layer s is the same since according to the available information individual nodes cannot be distinguished based on their connectivity. Further, the transition probabilities between any two nodes $i \in G_s$ and $j \in G_t$ are the same for any two fixed layers s and t. Therefore, all block transition matrices $\hat{\mathbf{T}}_{st}$ contain the same value at each entry. Hence we have

$$\hat{\mathbf{T}}_{st} = \left\{ \frac{m_{st}}{n_t \left(\sum_k m_{sk}\right)} \right\}_{ij}, \quad i \in \{1, \ldots, n_s\}, \quad j \in \{1, \ldots, n_t\}. \tag{6.9}$$

Now, using Eq.(6.5) we can construct an aggregated supra-transition matrix $\mathfrak{T}$ with entries

$$\alpha_{st} = \frac{m_{st}}{\sum_k m_{sk}}. \tag{6.10}$$

The aggregated supra-transition matrix $\mathfrak{T}$ describes the macro behavior of the multi-layer network ignoring the detailed topology of the inter-links and the intra-links of all single layers. Since $\mathfrak{T}$ depends on a mean-field approach it only captures probabilistic assumptions of the ensemble $\mathcal{E}(\vec{n}, \mathbf{M})$. Thus, the spectrum of the mean-field supra-transition matrix $\hat{\mathbf{T}}$ can be calculated by

$$Spec(\hat{\mathbf{T}}) = Spec(\mathfrak{T}) \cup \left(\bigcup_{s=1}^{L} \cup_{i=1}^{n_s - 1} \{0\} \right). \tag{6.11}$$

To clarify the situation, let us briefly discuss the simple case of a network $\mathbf{G}$ that contains only two layers G_1 and G_2, for which we get

$$\mathfrak{T} = \begin{pmatrix} 1 - \alpha_{12} & \alpha_{12} \\ \alpha_{21} & 1 - \alpha_{21} \end{pmatrix}. \tag{6.12}$$

Hence, for the mean-field matrix of a two-layered network we obtain

$$Spec(\hat{\mathbf{T}}) = \{1, 1 - \alpha_{12} - \alpha_{21}, \underbrace{0, \ldots, 0}_{|n|-2 \text{ times}}\}. \tag{6.13}$$

These results are remarkable, since the layer-aggregated transition matrix captures the same relevant eigenvalues as the mean-field transition matrix. So, for the case of a diffusion process in two layers the eigenvalue of interest is $\lambda_2(\hat{\mathbf{T}}) = 1 - \alpha_{12} - \alpha_{21}$. However, so far we only considered the general case where we can only use the densities of inter-links and intra-links of all single layers. In the following two sections we will investigate cases where we may have some additional information about either the inter-link topology between all single layers or the intra-link topology of all single layers. For simplicity, we will restrict ourselves to the two layer case but, as shown in the appendix, our results can be generalized to multiple layers.

6.3.2 Case II: unknown inter-connectivity

For this case we assume full knowledge of the intra-link topology, i.e. we know exactly which nodes are connected in all of the single layers. But while we know the number of links between the layers we do not know how the layers are connected, i.e. we do not know the inter-link topology. With respect to the general case discussed previously, here we have more information which is expected to improve the estimates of the ensemble average.

More precisely, we consider a two-layer network with unknown inter-link structure denoted by $E_I(\mathbf{G})$, but with a given amount of m interconnecting links which connect the networks G_1 and G_2. This means that the diagonal blocks $\mathbf{A}_1$ and $\mathbf{A}_2$ of the supra-adjacency matrix are given, but the off-diagonal blocks $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ can take any form such that they have exactly m entries different from zero. Since there are no further constraints on the ensemble, any random link configuration that consists of m inter-links has the same probability to occur. Therefore, we define the mean-field supra-adjacency blocks that correspond to the inter-links, $\hat{\mathbf{A}}_{12}$ and $\hat{\mathbf{A}}_{21}$, to have the same value $\frac{m}{n_1 n_2}$ in each entry.

For the supra-transition matrix we have to row-normalize $\mathbf{A}_1$ with $\hat{\mathbf{A}}_{12}$ and $\hat{\mathbf{A}}_{21}$ with $\mathbf{A}_2$. The row sums of $\hat{\mathbf{A}}_{12}$ are all equal to m/n_1 and the row sums of $\hat{\mathbf{A}}_{21}$ are all equal to m/n_2, while the row sums of $\mathbf{A}_1$ and $\mathbf{A}_2$ correspond to the individual degrees of the nodes in G_1 and G_2 respectively. Thus, we use the mean degree $\hat{d}_1$ of G_1 and $\hat{d}_2$ of G_2 in order to obtain the row-normalized transition matrix $\hat{\mathbf{T}}$, and to define the following factors

$$\alpha_1 = \frac{n_1\hat{d}_1}{n_1\hat{d}_1 + m}, \quad \alpha_2 = \frac{n_2\hat{d}_2}{n_2\hat{d}_2 + m}, \quad \alpha_{12} = \frac{m}{n_1\hat{d}_1 + m}, \quad \alpha_{21} = \frac{m}{n_2\hat{d}_2 + m}\,. \tag{6.14}$$

Note that $\alpha_1 + \alpha_{12} = 1$ and $\alpha_2 + \alpha_{21} = 1$.

Accordingly we define the mean transition blocks of $\mathbf{T}_{12}$ and $\mathbf{T}_{21}$.

$$\hat{\mathbf{T}}_{12} = \left\{ \frac{m}{n_2(n_1\hat{d}_1 + m)} \right\}_{ij} \quad \text{for} \quad i \in \{1, \ldots, n_1\}, j \in \{1, \ldots, n_2\} \tag{6.15}$$

$$\hat{\mathbf{T}}_{21} = \left\{ \frac{m}{n_1(n_2\hat{d}_2 + m)} \right\}_{ij} \quad \text{for} \quad i \in \{1, \ldots, n_2\}, j \in \{1, \ldots, n_1\}\,. \tag{6.16}$$

This means that each of the off-diagonal block matrices that correspond to the mean-field inter-link structures have the same value at each matrix element, and the diagonal blocks are just rescaled transition matrices of $\mathbf{A}_1$ and $\mathbf{A}_2$,

$$\hat{\mathbf{T}}_1 = (1 - \alpha_{12})\, T(\mathbf{A}_1), \quad \hat{\mathbf{T}}_2 = (1 - \alpha_{21})\, T(\mathbf{A}_2)\,, \tag{6.17}$$

where $T(\mathbf{M})$ is the row-normalized version of matrix $\mathbf{M}$. We denote the supra-transition matrix with the blocks constructed as described before by $\hat{\mathbf{T}}$,

$$\hat{\mathbf{T}} = \left(\begin{array}{c|c} \hat{\mathbf{T}}_1 & \hat{\mathbf{T}}_{12} \\ \hline \hat{\mathbf{T}}_{21} & \hat{\mathbf{T}}_2 \end{array} \right)\,. \tag{6.18}$$

This mean-field matrix has some special properties. First of all, the eigenvalues of $\hat{\mathbf{T}}_1$ and $\hat{\mathbf{T}}_2$ are also eigenvalues of $\hat{\mathbf{T}}$. Further, the multi-layer aggregation of $\hat{\mathbf{T}}$ is given by

$$\mathfrak{T} = \begin{pmatrix} \alpha_1 & \alpha_{12} \\ \alpha_{21} & \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 - \alpha_{12} & \alpha_{12} \\ \alpha_{21} & 1 - \alpha_{21} \end{pmatrix}\,, \tag{6.19}$$

so, the second-largest eigenvalue of $\mathfrak{T}$ is given by $\lambda_2 = 1 - \alpha_{12} - \alpha_{21}$.

The second-largest eigenvalues of $\hat{\mathbf{T}}_1$ is equal to $(1-\alpha_{12})\lambda_2^1$ and of $\hat{\mathbf{T}}_2$ is equal to $(1-\alpha_{21})\lambda_2^2$, where $\lambda_2^1 = \lambda_2(T(\mathbf{A}_1))$ and $\lambda_2^2 = \lambda_2(T(\mathbf{A}_2))$. Therefore the second largest eigenvalue of $\hat{\mathbf{T}}$, denoted by $\lambda_2(\hat{\mathbf{T}})$, fulfills the following condition (See Prop. 2 in the Appendix B.1 for more details)

$$\lambda_2(\hat{\mathbf{T}}) = \max\left(1 - \alpha_{12} - \alpha_{21}, (1 - \alpha_{12})\lambda_2^1, (1 - \alpha_{21})\lambda_2^2\right)\,. \tag{6.20}$$

We would like to remind the reader that an eigenvalue λ_2 close to one implies slow convergence and λ_2 close to zero fast convergence. From the above equation we can see that as long as $\lambda_2 = 1 - \alpha_{12} - \alpha_{21}$ is maximal the inter-links are the limiting factor of the convergence in the multi-layer network. This means that due to the inter-link topology the random walk diffusion is slowed down, and the influence of the intra-layer topologies

is marginal to the process.

When either the term of λ_2^1 or λ_2^2 is maximal then the diffusion is limited by the single layer G_1 or G_2, and the additional information provided by the intra-layer topologies becomes relevant as it affects the diffusion process. Note that the change between λ_2 and either λ_2^1 or λ_2^2 being maximal is related to the transitions pointed out in Ref. [55, 131].

This behavior is shown in Figure 6.1 for the mean-field matrix of two interconnected networks. The figure shows the second largest eigenvalues of $\hat{\mathbf{T}}, \mathfrak{T}$ and the sparsest layer $\mathbf{T}_1$ for different amount of inter-links. When only a few inter-links are present the interconnectivity between layers slows the process down, as it is expected. When we increase the amount of inter-links, we can reach the convergence rate of single layers, which is the point where the single layers slow down the process. However, with an increasing amount of inter-links the single layers lose their importance and the process is again slowed down by the inter-links. This happens because a very large amount of inter-links force the random walker to switch between layers with increasing probability, thus, preventing diffusion to reach the whole layer. To conclude, the mean-field transition matrix $\hat{\mathbf{T}}$ is a better estimation than $\mathfrak{T}$ in intermediate numbers of inter-links, which for our systems is in the region of approximately 550 to 1800 inter-links. Otherwise, the information about the link densities as captured in $\mathfrak{T}$ is enough to approximate the second-largest eigenvalue of $\hat{\mathbf{T}}$, and thus the speed of diffusion.

In general the spectrum of a mean-field matrix $\hat{\mathbf{T}}$ with unknown inter-link topology is given by

$$Spec(\hat{\mathbf{T}}) = \{1, \lambda_2, \ldots, \lambda_n\} \cup \left(\bigcup_{s=1}^{n} Spec(\hat{\mathbf{T}}_s) \setminus \lambda_1(\hat{\mathbf{T}}_s) \right) , \tag{6.21}$$

or

$$Spec(\hat{\mathbf{T}}) = Spec(\mathfrak{T}) \cup \left(\bigcup_{s=1}^{n} Spec(\hat{\mathbf{T}}_s) \setminus \lambda_1(\hat{\mathbf{T}}_s) \right) , \tag{6.22}$$

where $\mathfrak{T}$ is the multi-layer aggregation of $\hat{\mathbf{T}}$ as described before (for details see Prop. 2 in the Appendix B.1). This decomposition of eigenvalues can also be useful for other network properties that depend on eigenvalues.

So far we provided an estimation based on the eigenvalues of a mean-field transition matrix $\hat{\mathbf{T}}$ that intends to approximate the ensemble average. In reality however, ensemble realizations of multi-layer networks that contain layer G_1 and G_2 can deviate from the mean-field estimation. This is shown in Figure 6.2 (a) where we plot the second-largest eigenvalues of $\hat{\mathbf{T}}$, $\mathfrak{T}$, and ensemble averages over 100 realizations of $\mathbf{T}$ against the number of inter-links between G_1 and G_2. As we can see, the magenta colored dashed line showing

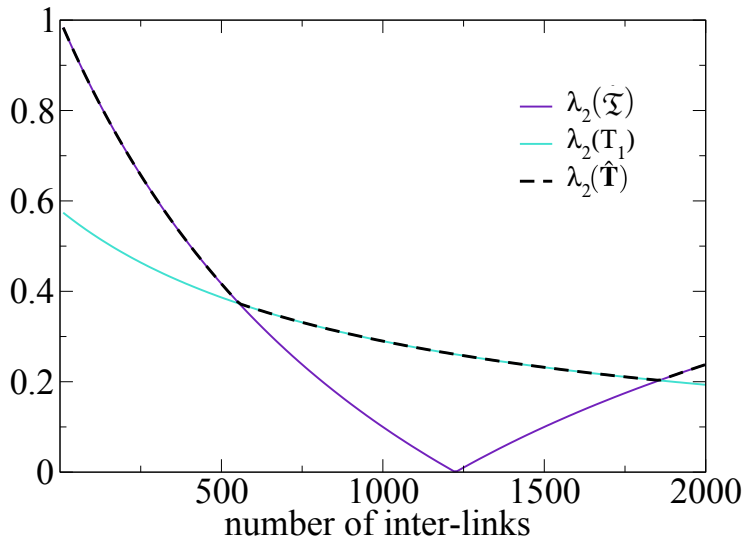


Figure 6.1: Eigenvalues of a mean-field approach of a two-layered network Layer 1 consists of an Erdös-Rényi network of 100 nodes and 500 links and Layer 2 consists of an Erdös-Rényi network of 100 nodes and 750 links. The x-axis indicates the amount of inter-links randomly added across the layers. The lines indicate the second-largest eigenvalue of: black dashed: the mean-field supra-transition matrix $\lambda_2(\hat{\mathbf{T}})$, violet: the layer-aggregated matrix $\lambda_2(\mathfrak{T})$, turquoise: the larger single layer eigenvalues of $\lambda_2(\mathbf{T}_1)$ and $\lambda_2(\mathbf{T}_2)$.

the mean-field approximation of $\mathfrak{T}$ is a good proxy for the diffusion dynamics in the region when inter-links dominate, which is the case for either sparse or very dense inter-link topologies. However, as shown by the cyan colored line, we can actually improve this approximation if we additionally consider the intra-links of all single layers.

There is a peak where the difference between the estimation and the ensemble averages $\Delta\lambda_2 = \lambda_2(\mathbf{T}) - \lambda_2(\hat{\mathbf{T}})$ reaches high values up to 0.225, as shown in Figure 6.2 (b). This happens, on one hand, due to the large degree of freedom that comes from the absence of intra-connectivity informations within the layers. On the other hand, the mean-field matrix assumes "full-connectivity" across layers, and even though this implies small weights for each single inter-link, it leads to a systematic bias towards overestimating the diffusion speed. Nevertheless, we would like to highlight that the multi-layer aggregation provides a quite accurate estimation of the diffusion speed in the regimes where inter-links limit diffusion.

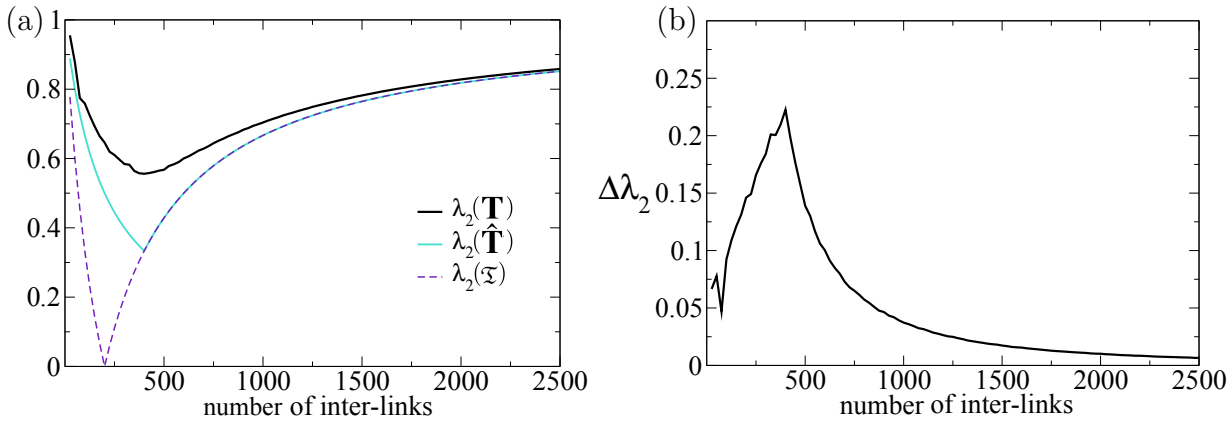


Figure 6.2: Mean-field approach of a two-layered network Layer 1 and 2 both consist of an Erdös-Rényi network of 50 nodes and 100 links but with different topologies. **a)** Eigenvalues of the multi-layer network. The x-axis indicates the amount of random inter-links added across layers. The lines indicate the second-largest eigenvalues of: black line: ensemble averages, turquoise line: mean-field estimate including intra-link topology, violet dashed line: mean-field only considering densities. **b)** Eigenvalue difference between ensemble average and mean-field estimation $\Delta\lambda_2 = \lambda_2(\mathbf{T}) - \lambda_2(\hat{\mathbf{T}})$.

6.3.3 Case III: unknown intra-connectivity

For this case we assume full knowledge of the inter-link topology, i.e. we know exactly how the layers are connected, but the intra-link topologies, i.e. how the nodes are connected within the single layers, are unknown. More precisely, we consider two interconnected layers G_1 and G_2 of a multi-layer network, and we fix the inter-links $E_I(\mathbf{G})$ in a bipartite network structure that connects nodes of G_1 to nodes of G_2. Since we have no information about the intra-link topologies of G_1 and G_2, we assume random connectivities within the layers, so that we only know the average degrees $\hat{d}_1$ and $\hat{d}_2$ of G_1 and G_2 respectively. This means that the off-diagonal blocks $\mathbf{A}_{12}^{\top} = \mathbf{A}_{21}$ of the supra-adjacency matrix are given, but the diagonal blocks $\mathbf{A}_1$ and $\mathbf{A}_2$ are unknown.

Because we only know the average degrees $\hat{d}_1$ and $\hat{d}_2$ of the layers, we can define mean-field versions of the adjacency matrices such that

$$\hat{\mathbf{A}}_1 = \left\{ \frac{\hat{d}_1}{n_1} \right\}_{ij} \quad \text{and} \quad \hat{\mathbf{A}}_2 = \left\{ \frac{\hat{d}_2}{n_2} \right\}_{ij} .$$

However, even though we know the topology of the inter-links, we do not know which nodes exactly are connected to each other. Hence we use the same approach as in Case II with

m equal to the amount of inter-links and the factors defined as in Eq.(6.14). Therefore we get the mean-field transition matrix $\hat{\mathbf{T}}$ consisting of the following block matrices,

$$\hat{\mathbf{T}}_1 = \left\{ \frac{\hat{d}_1}{n_1\hat{d}_1 + m} \right\}_{ij} \quad \text{for} \quad i \in \{1, \dots, n_1\}, j \in \{1, \dots, n_2\} \tag{6.23}$$

$$\hat{\mathbf{T}}_2 = \left\{ \frac{\hat{d}_2}{n_2\hat{d}_1 + m} \right\}_{ij} \quad \text{for} \quad i \in \{1, \dots, n_2\}, j \in \{1, \dots, n_1\}\,. \tag{6.24}$$

The off-diagonal blocks are just rescaled transition matrices of $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$,

$$\hat{\mathbf{T}}_{12} = \alpha_{12} T(\mathbf{A}_{12}), \quad \hat{\mathbf{T}}_{21} = \alpha_{21} T(\mathbf{A}_{21})\,. \tag{6.25}$$

However, this time we are not able to compute exactly the single layer eigenvalues λ_1^1 and λ_2^2, as it was the case in Case II. In particular, depending on the ensemble constraints we could only compute an average eigenvalue $\hat{\lambda_2}$ for a single layer. Therefore, we can use the following maximization term

$$\lambda_2(\hat{\mathbf{T}}) = \max\left(1 - \alpha_{12} - \alpha_{21}, (1-\alpha_{12})\hat{\lambda_2^1}, (1-\alpha_{21})\hat{\lambda_2^2}\right)\,, \tag{6.26}$$

which is has the same form as in Case II (see Eq.(6.20)). Here again, as long as $\lambda_2 = 1 - \alpha_{12} - \alpha_{21}$ is maximal the inter-links are the limiting factor of diffusion in the multi-layer network, which means that due to the inter-link topology the random walk diffusion is slowed down, and the influence of the intra-layer topologies is marginal to the process. On the other hand, when either the average term of $\hat{\lambda_2^1}$ or $\hat{\lambda_2^2}$ is maximal then the diffusion is limited by the single layer G_1 or G_2, and the additional information provided by the intra-layer topologies becomes relevant as it affects the diffusion process.

In Figure 6.3(a), starting with initially empty intra-networks[1], we plot the second largest eigenvalues of $\mathfrak{T}$, $\hat{\mathbf{T}}$, and the ensemble average of 100 realizations of $\mathbf{T}$ against the number of intra-links that are simultaneously and randomly added in both layers. We observe that the general behavior is similar to Figure 6.2. Thus, the multi-layer aggregation plotted in magenta approximates well the regions where the inter-links are the relevant factor, which is for very sparse and increasingly dense intra-links densities. The difference between the mean-field and the ensemble average $\Delta\lambda_2 = \lambda_2(\mathbf{T}) - \lambda_2(\hat{\mathbf{T}})$ as seen in Figure 6.3(b) again rises up to a peak of about 0.225.

Our analysis shows that there is some form of symmetry in knowing the degree of the

[1]Note that even though the intra-layer networks are empty initially, there is a number of inter-layer links which provide connectivity across the layers, similar to a bipartite network.

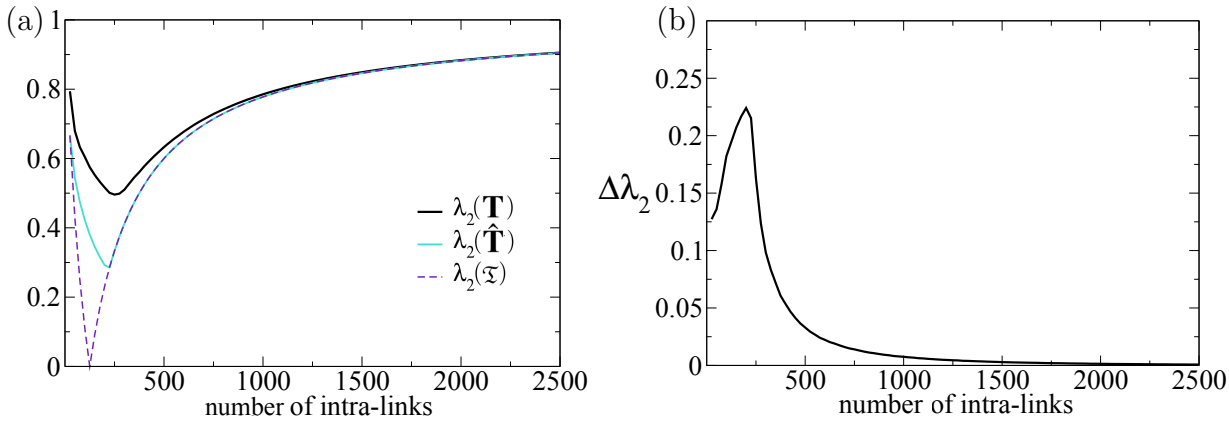


Figure 6.3: Mean-field approach of a two-layered network a) Eigenvalues of a two-layered network with 250 inter-links. Layer 1 and 2 both consist of 50 nodes but no links. The x-axis indicates the amount of links intra-links that are simultaneously added to both layers. The lines indicate the second-largest eigenvalue of: black line: ensemble averages, turquoise line: mean-field estimate including inter-links, dashed violet line: mean-field only considering link densities. **b)** Eigenvalue Difference between ensemble average and mean-field approach $\Delta\lambda_2 = \lambda_2(\mathbf{T}) - \lambda_2(\hat{\mathbf{T}})$.

nodes in the single layers, but not knowing how they are connected to nodes in other layers and to knowing the inter-links between layers, but not the degree of their adjacent nodes. Even though the ensembles generated from these two constraints can be much different, the relevance of inter-links or intra-links of all single layers to a diffusive process is comparable for both cases.

6.4 Conclusion

In this chapter, we showed how an ensemble perspective can be applied to multi-layer networks in order to address realistic scenarios when only limited information is available. More precisely, we focused on a diffusion process that runs on the multi-layer network and its relation to the spectrum of the supra-transition matrix. We have shown that the convergence rate of the diffusion process is limited by either the inter-links or intra-links of the single layers and we identified for which relation of inter-link compared to intra-link densities it is sufficient to only consider transitions across layers, instead of the full information on all individual nodes. This implies that we do not always need perfect information to make statements about a multi-layer network because, under certain conditions, we are still able to make analytical statements about the network only using

partial information. In realistic situations data can be an issue either due to constraints or due to their vast amounts. In such cases, even though an exact analysis is impossible, we may still derive useful conclusions about processes that depend on the network spectrum (like diffusion and synchronization) using only aggregated statistics.

For our study we assumed the simplest case of random networks, therefore exploring other ways to couple the network layers or including link-weights and directed links and testing their influence on our results is up to future investigation.

Chapter 7

Scientometrics: Social Influence on Citations

Summary

In this chapter we investigate how citations of scientific publications are influenced by social contacts among scientist. We use a multi-layer perspective to distinguish two relations between authors of scientific articles. On the one hand we use co-authorship as proxy of social contacts and citations as representation of the acknowledgment of particular works. We measure for how many citations we can find a previous co-authorship involving authors of the citing and cited article. Using this measure we quantify how strong the social bias is for particular subfields in physics. The prevalence of this bias is an indication of a social filtering mechanism that could hamper the scientific knowledge transfer. Incorporating the social bias of co-authorships we further provide a generative model for citations. Comparing the expected citations to the actual data we discuss citation based ranking schemes. Finally, we provide an indicator that allows to asses how strongly citations are correlated with co-authorships. Our analysis provides a multi-layer perspective on a real complex system and it allows to disentangle multiple effects that influence the formation of citation links.

The work presented in this chapter is based on a research project conceived by NW, IS, AG and FS. The results are planned to be published in the field of scientometrics. NW analyzed the data, designed the model, coded the algorithms, performed the simulations, did the plots and figures and wrote the main text.

7.1 Introduction

The interaction and relations between scientist and there publications create a complex system and the amount of knowledge produced is increasing every moment. It is not possible anymore for an individual scientist to be aware of all works that are part of their discipline. Hence, it gets harder to filter the huge amount of information for the content that is relevant for a particular scientist. In this sense, relevance is attributed to works that provide the most impact in the scientific community and therefore are most likely of higher quality. To deal with the filtering problem it is needed that scientific publications can be quantified in such a manner that allows to indicate the most qualitative results or findings.

A first attempt to capture the relation of scientific works was done by Eugen Garfield who created the *Science Citation Index (SCI)* [50]. The idea of this index is to create a bibliographic database that lists the citations between scientific publications of academic journals. This allows to identify the time when a work was published and to which other publications it refers to. The relational data of publications based on citations gives rise to a system that can be analyzed by network metrics and measures. One of them is the citation count of each publication which expresses how many other works acknowledge the findings. Also more elaborate measures can be applied to study the relations between scientific publications. Several studies were concerned with implications and interpretations of these measures and therefore laid the foundation of a new research field.

The quantitative aspects of the science of science is commonly abbreviated by *scientometrics*. This new discipline became popular in 1978 when a overview work [36] presented this novel perspective. In the same year the journal *Scientometrics* was launched that fostered such research. The field today is not only based on quantitative science but also has broad relations to philosophy and sociology.

A lot of attention is paid to *indicators* that allow to classify or rank scientific works according to several criteria. A straightforward indicator is the mere amount of citations that an article or scientist receives. The idea behind the citation count is the fact that works that are acknowledged a lot should also contain more significant content that was of importance for a lot of works. This measure is often aggregated to individual scientist by summing up all of the citations that their papers received. However, usually a lot of properties affecting the citation count of scientists like the amount of articles they published or their seniority are neglected. To deal with some of these issues, more advanced measure such as the h-index [65] were introduced and are still being explored and refined [35]. There

are various indicators that are used to measure importance or quality of publications, but it is not always clear what exactly they capture [168].

However, there can be several reasons why an author decides to cite a particular publication. The motivation can, but does not have to, be relevance or scientific quality. The various reasons for a citation were studied extensively [20, 21, 95, 96, 167]. One aspect that was found to have significant impact is the collaboration between scientists [41, 120, 169]. There are several reasons why collaborations can influence the citation behavior. One of them is the issue mentioned above, that there are increasingly more works and it is not possible to know all of them. Nevertheless, scientists are normally more aware of the works of their collaborators and therefore also more likely to cite them. In particular, it was found that scientific success in terms of citations is correlated with co-authorships [98, 103, 143]. Co-authorships represent a possible way of collaboration and usually require a social interaction between the authors.

In this chapter we focus on the social effect of collaborations respectively co-authorships onto citations. We hypothesize that due to the increasing amount of publications over time, scientists rely more and more on their social network. This is related to the so-called *filter bubble* [121] that describes the effect of personified recommender schemes employed by Google and Facebook. This personalization leads to an individual filtering of information content since users get prominently confronted with things that are in line with their previous online history and therefore new content is suppressed. The same effect likely is present in the scientific landscape. Due to the huge amount of publications scientists are more exposed to works by previous collaborators, therefore diverting the attention from possibly relevant works done by unknown people.

In the following we first focus on social effects in citation behavior and use the acquired insights to discuss ranking schemes. More precisely, we analyze for each citation if there was previously a co-authorship between any author of the citing publication and any author of the cited publication. We interpret a common co-authorship as an act of collaboration and therefore a social interaction between the authors which makes them more aware of each other's work. Showing that there is a significant overlap of citations and collaborations, we use this correlation to create a random citation model. Based on this model, we can then estimate which amount of citations we would expect at random.

Our study is based on a bibliographic dataset that we analyze from a multi-layer perspective. Here, the multi-layer network will on one hand contain layers with different node types, and on the other hand, span multi-dimensional relations between the same type of nodes. More precisely, we combine the layer of citations between articles with the layer of co-authorships between scientists. By further connecting these layers with links

between publications and their corresponding authors we are able to project and correlate the properties of single layers across the whole network. The cross layer relations will be crucial in assessing correlations in the data that would not be possible from a single layer perspective.

7.2 Multi-layer perspective

Before we present our investigations, we first introduce the methodology used throughout this chapter. To analyze the citation topology between scientists according to additional parameters we have to represent them in an appropriate way. Citations indicate a directed relation between scientists and therefore give rise to a citation network. We want to investigate the influence of collaboration on citations. Collaborations, or more precisely co-authorships, can be interpreted as social relations between scientists. Hence, we naturally deal with different link types in the network of scientists, which calls for a representation as a multi-layer network.

In this section we discuss how we can apply the multi-layer network approach to a data set of scientific publications. Chapter 6 dealt with multi-layer networks that in respect to a diffusion process consisted of the same type of links and nodes. However, here we investigate a multi-layer network that includes different link types. An extended representation of the data set also consists of different types of nodes. Therefore the perspective that we take in this chapter is a complex multi-layer network comprising multiple dimensions of information.

7.2.1 Bibliographic data set

We base our analysis on a bibliographic data set. From this data set all needed information about citations and collaborations of scientist is extracted. Therefore, we do not use any additional knowledge and deal with a clearly circumscribed system.

In the following we describe the application of a multi-layer network approach to a data set of scientific publications. The data that we investigate is provided by the *American Physical Society (APS)* [1]. It contains bibliographic information on over $450,000$ articles published in APS journals between 1893 and the end of 2009. The data consist of citing article pairs and meta information on all articles. For our analysis we use the DOI, authors, affiliations, *PACS* number and the printing date.

PACS The Physics and Astronomy Classification Scheme® (PACS) was developed by the American Institute of Physics (AIP) and is used since 1975 to identify fields and subfields in physics. According to it subjects, each article can be tagged by one or several PACS numbers that it belongs to. As of 2010 there are 10 main fields that cover the following topics:

00 General
10 The Physics of Elementary Particles and Fields
20 Nuclear Physics
30 Atomic and Molecular Physics
40 Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics
50 Physics of Gases, Plasmas, and Electric Discharges
60 Condensed Matter: Structural, Mechanical and Thermal Properties
70 Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties
80 Interdisciplinary Physics and Related Areas of Science and Technology
90 Geophysics, Astronomy, and Astrophysics

We use the PACS classification scheme to divide the data set in to subfields. By this, we avoid part of the heterogeneity present in different fields that is due to different citation or collaboration behavior. Further, by only comparing publications belonging to the same PACS class we assure that they are more similar and relevant to each other than two randomly chosen publications from the whole spectrum of the APS journals.

Article citation network The publications of any APS journals are called articles. If an article wants to acknowledge the findings of other articles it can list them as reference. Such a reference leads to a citation between two articles and by this naturally forms a citation link. Each citation link is directed from the *citing* article to the *cited* article. The citation gets time-stamped with the printing date of the citing article. In general, citations between articles are directed and unweighted. I.e., articles can not mutually cite each other and there is only one citation counted between any pair of articles independent on how many times the cited article is referenced to in the citing article. All citation links together form the *article citation network*. Since the network is directed and the links are time-stamped the network contains no loops. The DOI is used as a unique identifier to label the nodes in the article network.

7.2.2 Projection and aggregation to article meta information

Using the meta information of each article allows to build projected networks. The meta information we primarily use are the *authors* and *affiliations*. The authors represent scientist that published in one of the APS journals and the affiliations correspond to the institutions or universities which are attributed to a specific author. To represent the correspondence of each author or affiliation to a certain article we form a link between them. These links are considered to be *correspondence links* and connect (i.e. indicate a realtion between) different kind of nodes. The correspondence links are undirected and unweighted and play the role of inter-layer links. This way, article nodes are connected with author nodes and affiliation nodes. Additionally, we can also build correspondence links between author nodes and affiliation nodes if needed. However, for our purpose it is enough to be able to link an article to all of its authors and affiliations.

Author co-authorship network We consider the author list of each article as unordered and do not distinguish "first-authors" or any other hierarchy of contribution. We regard the co-authoring of an article as a social mechanism. Therefore we link each pair of co-authors of an article by a *co-authorship link*. This implies that each article gives rise to a clique of authors that are all linked to each other. See Figure 7.1 (a) for a illustration. Since the authors are considered to collaborate mutually with each other the links are undirected. We assign each co-authorship link a time-stamp that is equal to the printing date of the corresponding article. Authors of multiple articles are part of multiple co-authorship cliques. All these cliques together create the *author co-authorship network*. Even though multiple co-authorships of the same authors on different articles could be counted, here we consider an unweighted network,

Author citation network Since each article is linked to all of its authors, each article citation gives rise to *author citations*. To obtain citations between authors, we *project* article citation to the author layer such that each author of the citing article will cite each author of the cited article. This yields a complete bipartite network with the independent set composed of authors from either the citing or cited article. See Figure 7.1 (b) for illustration. Through this process an author from the cited article can receive multiple citations from authors of the same citing article. Depending on the situation it is therefore advisable to use fractional counting for the author citation links. Hence the weights of the incoming author citation links are divided by the amount of authors that worked on the citing article. This implies that an author receives author citations that sum up to at most one for a single article citation that is projected to the author layer. Performing such

a projection for all article citation links results in a weighted and directed *author citation network*.

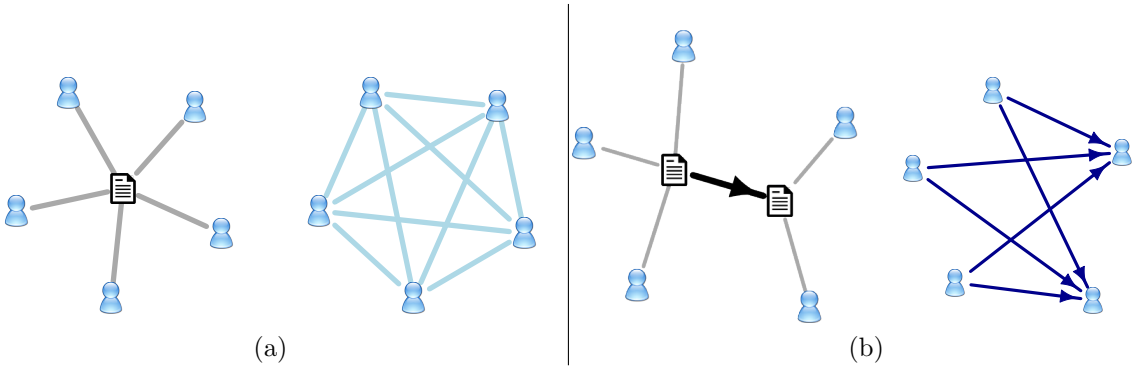


Figure 7.1: Author Collaboration and Citation Network (a) Construction of the author co-authorship network from the author article correspondence. All authors of an article are linked to each other with an unweighted co-authorship link. (b) Construction of the author citations network from article citations. All authors of the citing article form an out-going citation link to all authors of the cited article. In case of fractional link counting, each author citation link in this case would have a weight equal to 1/3.

Affiliation co-authorship and citation network Each author in the data set can be associated with an affiliation. This allows to attribute each article to the affiliations of its authors. Note however, that we do not track movements of authors between affiliations and therefore only their most recent affiliation is considered. Therefore, there may be inaccurate links between articles and affiliations when the actual credit of an article should belong to an institution in which an author previously worked. Nevertheless, we regard this inaccuracy as negligible from a macro perspective. Therefore we can construct *affiliation citation* and *co-authorship networks* in the same way as the author networks. The citations and co-authorships between affiliations are just an aggregation of the corresponding nodes and links in the author networks. However, also in case of affiliations a fractional counting is applied such that a single article citation only accounts for incoming affiliation citations that sum up to 1.

7.2.3 Multiplex network of citations and co-authorships

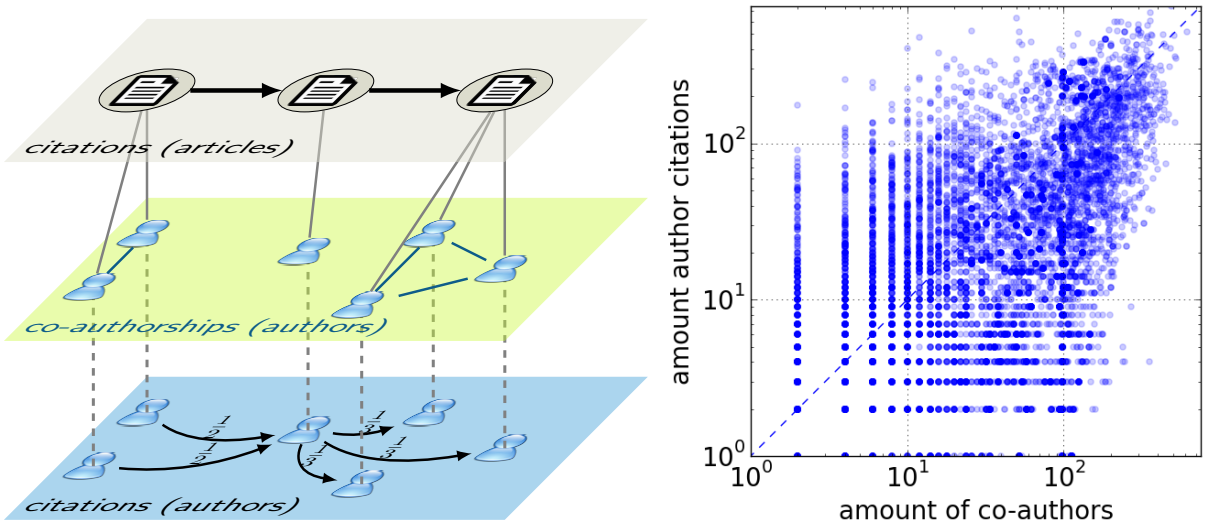
The co-authorships and citation layers of authors and affiliations allow to construct a multiplex network, see Figure 7.2 (a). In both layers we have the same kind of nodes, which are either authors or affiliations. This way we can compare two types of links:

citation links and co-authorship links. There is already a lot of information included in this multiplex. However, one should be aware of some technical details before interpreting the link topologies. While citations are directed from citing to cited nodes the co-authorship links are undirected. This means that we should differentiate in- and out-degrees when investigating properties of nodes. Further, one has to consider whether the weights of links that were projected from the article citation network are normalized or not. The weight of incoming citations can be adjusted by the amount of authors that worked on the same article. The same can also be done for co-authorships which can also be re-weighted according to the amount of authors that worked on the same article.

To get a first impression on how co-authorships and citations among authors are related, we first focus on *distinct* relations. By the amount of distinct co-authorships we mean with how many different authors a given scientist collaborated without considering how many times a collaboration took place with a particular co-author. This can easily be analyzed based on the degree in an unweighted author co-authorship network. In the same way we define the amount of distinct author citations that is equal to the amount of different authors that cite any article for a given scientist. This quantity can be derived from the in-degree of an unweighted author citation network.

As an example, we investigate articles classified by PACS 20 for an observation period of 10 years from 2000 to 2010. For all scientists that published an article during this period, we evaluate the amount of distinct co-authors they worked with and the amount of distinct authors that cited at least one of their articles. We represent the result in a scatter plot shown in Figure 7.2 (b). Note that the plot has a logarithmic scale. We can see that there is a tendency for scientists with a lot of co-authors to also receive a lot of citations whereas for low citation and co-author counts there is larger dispersion. In other word more citations come along with more co-authorships and vice versa. The correlation of citation and co-author counts can be due to many reasons. One is the pure amount of publications that were published by a single scientist. A scientist with a high co-author count usually also published a lot of articles. Each article has a chance to receive citations therefore also increasing the total citation count of the scientist publishing them. The data reveals also some outliers. On the one hand there some scientists that only collaborated with a few scientists but still have a high citation count, on the other hand there a some scientists that worked with a lot of co-authors but receive comparably few citations. Here, we only showed PACS class 20 for a certain time period as an example, however also the other PACS classes and time windows exhibit similar correlations.

With this brief investigation we just wanted to illustrate how co-authorships and citations between authors are correlated. We do not intend to argue why a particular author has



(a) Multi-layer Network

(b) Scatter Plot

Figure 7.2: Scientometric Multi-Layer Network (a) Top: article citation network, middle: author co-authorship network, bottom: author citation network. Gray solid lines indicate correspondence links, gray dashed lines connect the same node in a multiplex layer. (b) Scatter plot where each point represent a scientist that is placed according to amount of distinct co-authors he worked with and the amount of distinct authors that cited one of his articles. The data is based on articles classified by PACS 20 for a time period from 2000 to 2010.

a certain co-authorship and citation count. There can by many reasons for this and we would need to analyze in detail each author individually. However, the multiplex perspective allows to put these different types of relation in perspective. Knowing how the the multi-layer network is composed we next focus on identifying the social influence on citation behavior.

7.3 Social influence on citations

As presented in the previous section, all the bibliographic information of citations and collaborations can be comprised into a multi-layer network. Further, we have seen that the amount of distinct citations and co-authorship is correlated. In this section we investigate the relationship between citation and co-authorship in more detail. By showing that co-authorships alter the citation behavior of scientist in a significant way, we intend to verify that there is indeed a social bias that affects citations.

As argued in the introduction we hypothesize that scientists are forced to apply filtering

mechanisms to keep track of works that are relevant to them. One strategy could be to rely more on the publications of collaborators that provide a source of information that can be more easily evaluated. Citations serve as indication of knowledge transfer from a certain publication, thus indicating that the citing authors acknowledged the work of the cited authors. To identify if such a citation could have been influenced by previous social contacts we need a measure that allows to relate them to each other. In this section we provide a global measure that tells how frequent citations are which can be related to social contacts. By comparing the real world data to an expectation based on a generative model we further assess how significant the findings are.

7.3.1 Citations matched by co-authorships

To study the potential social mechanism behind citations we need to link them to co-authorships. We intend to identify citations between articles where at least one author of the cited article previously collaborated with at least one author of the citing article. As discussed before each article can be linked to all of its authors. Let us define the set of authors $Auth(p)$ for any article p. It implies that there exists an article-author link between p and all authors $a_i \in Auth(p)$. Consider an article citation $cit_{art} : p_1 \to p_2$ where p_1 is the citing article and p_2 the cited article. For convenience we denote the set of all article citation links by Cit_{art} and the set of all author co-authorship links by Co_{auth}, hence $cit_{art} \in Cit_{art}$. For articles p_1 and p_2 we can create the set of authors $Auth(p_1)$ and $Auth(p_2)$ and further construct the set of *matched author* collaborations,

$$MatchedAuth(cit_{art} = (p_1, p_2)) = \{(a_i, a_j) \in Co_{auth} | a_i \in Auth(p_1), a_j \in Auth(p_2)\} . \tag{7.1}$$

Hence, $MatchedAuth(cit_{art})$ contains all co-authorships links that connect authors from p_1 and p_2. However, to derive any social influence or causality we have to consider the time-stamps of the co-authorship links between authors and the citation links of articles. Consider the function $t(\cdot)$ that returns the time-stamp of any citation or co-authorship link. Given an article citation cit_{art}, we only consider two matched authors if they had at least one co-authorship link that happened *before* the printing date of the citing article. Summarizing, we define the boolean function $matched(\cdot)$ from an article citations to either 0 or 1 as follows,

$$matched(cit_{art}) = H(|\{co_{auth} \in MatchedAuth(cit_{art}) | t(co_{auth}) < t(cit_{art})\}|) , \tag{7.2}$$

where $H(\cdot)$ is the Heaviside function, i.e. $H(x) = 1$ for $x > 0$ and otherwise $H(x) = 0$. Note that in our analysis we do not consider how many matched author pairs exist for a

given citation. Therefore, we apply the Heaviside function to constraint weights to either 0 or 1. However, the framework could be extended to weighted links by also counting the number of co-authorships.

Based on this function we construct the *matched article network*. We evaluate $matched(\cdot)$ for each possible citation link in the article network and if it returns 1 we include the link in the matched article network. Note that citations are only possible if the cited article was printed before the citing article. By construction the matched article network is a multiplex layer to the article citation network, see the illustration in Figure 7.3. It serves as a reference to check for all the matched citations that are possible. For each citation that is present in the article citation layer we can check if the corresponding link in the matched article layer is present as well. If the citation link is present in both layers it is considered to be a *matched citation*.

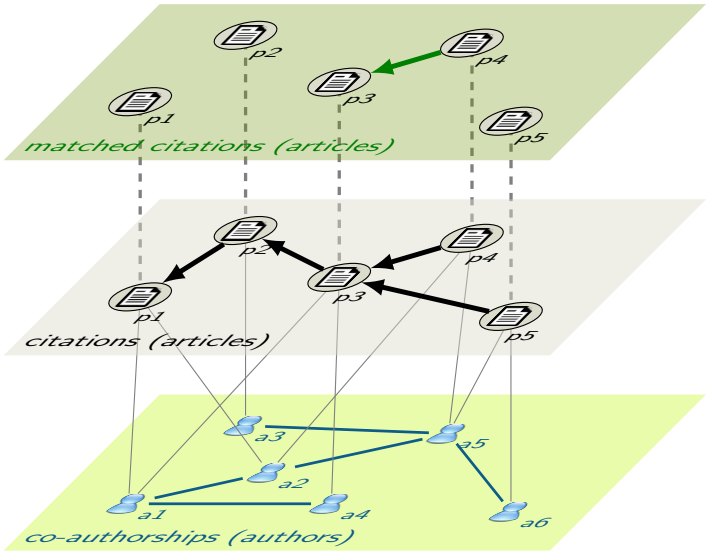


Figure 7.3: Concept of matched citations Top: matched article citation network, middle: article citation network, bottom: author co-authorship network. Gray solid lines indicate correspondence links, gray dashed lines connect the same node in a multiplex layer. There is one matched citation between p_4 and p_3 that is due to the co-authorship of a_1 and a_2.

The example illustrated in Figure 7.3 contains only one matched citation (p_4, p_3). Article p_3 is affiliated to authors a_1 and a_4 and article p_4 is affiliated to authors a_2 and a_5. Hence, we have that $Auth(p_3) = \{a_1, a_4\}$ and $Auth(p_4) = \{a_2, a_5\}$. Further, we can see that article p_1 is affiliated to the authors a_1 and a_2 which implies $Auth(p_1) = \{a_1, a_2\}$. Even though, the time-stamps are omitted in the figure, we can infer that p_1 was published before p_4 by the directionality of the citation links. Therefore, we conclude that $matched((p_4, p_3)) = 1$

which implies that (p_4, p_3) is a matched citation.

Give an article citation network Art_{cit}, we can compute the *fraction of matched article citations*. I.e. the number of citations in the article layer, that is also present in the matched article layer, compared to all citations in the article layer. We denote this measure by $FMC(\cdot)$,

$$FMC(Art_{cit}) = \frac{\sum_{e \in Cit_{art}} matched(e)}{|Cit_{art}|} \,. \tag{7.3}$$

The fraction of matched citation measure FMC tells us how many citations are based on previous co-authorships of the involved authors. We analyze this fraction throughout different data sets and time-windows over 20 years. To assure that the considered articles are more or less comparable we restrict our investigation to a 5 years time-window. A period of 5 years was used in previous works on scientometric data as a suitable time span that is representative for the acquired citations [110, 143]. This means that we only take into account the articles that were published during this time-window. We also only consider citations among this subset of articles. However, to identify if a given citation is matched or unmatched we consider the whole past of co-authorships ranging back until the start year of the data set. As mentioned before self-citation are neglected to not interfere with our analysis. To have more coherent groups according to the topic we investigate each of the main PACS classes individually. In Figure 7.4 the fraction of matched citations FCM is recorded on the y-axis for starting years on the x-axis. Each year indicates the starting point of a time-window that spans over the 5 following years.

We can observe that there is in general an upwards trend in the fraction of matched citations. However, the absolute size of FMC differs significantly for different PACS. We reach values even exceeding 20% for PACS class 20 and 50, while most of the other PACS classes reach about 10% of matched citations for the last time-window. Further plots showing the amount of articles published, the amount of publishing authors and the amount of total citations can be found in the Appendix B.2. We can summarize that all of these quantities do also show a general upwards trend.

The bottom line from this investigation is that matched citations got increasingly more frequent during the last two decades, even though the amount of articles and authors that are active increased as well which should allow for more diversity. There could be several reasons for the increase of matched citations. For example one could argue that a citation between the same pair of authors a_1 and a_2 in 2005 is more likely to be a matched citation as it was in 1990. Mainly due to the fact that we consider the full past of collaborations and therefore until 2005 there could be plenty of time to establish a co-authorship on a common article. However, there are of course also new authors entering the systems and

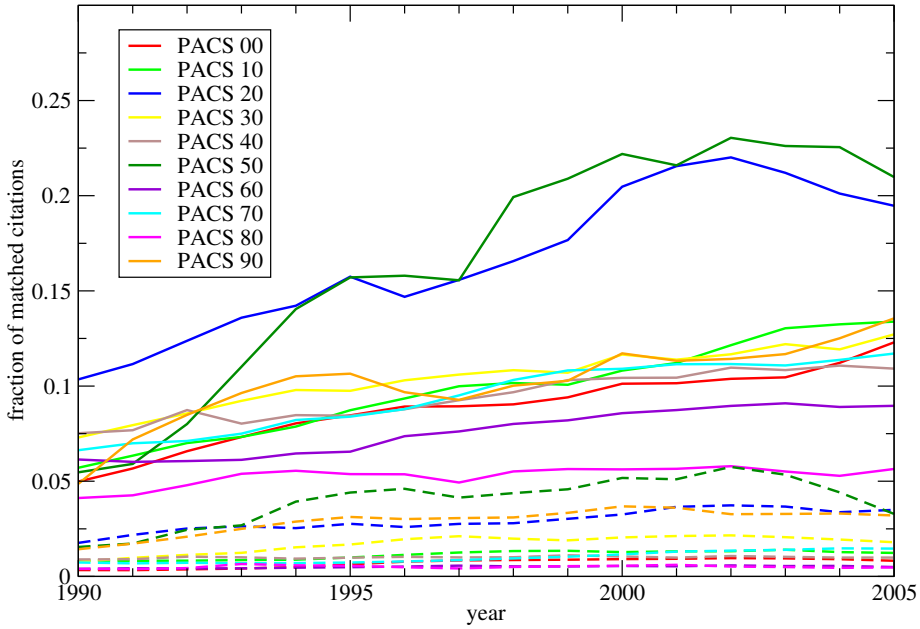


Figure 7.4: Fraction of matched citations FMC in the article citation network for sliding time-windows of 5 years The x-axis indicates the starting year of a time-window of 5 years. Only articles published during this period are considered for the analysis. The main PACS classes are represented by different colors. The solid lines represent the value of FMC for the actual data and the dashed line represent an average FMC value based on 100 simulations of the stochastic generative model $\mathcal{M}$.

older ones drop out. The FMC measure is a global quantity over all article citations and therefore does not directly account for such kind of changes in the topology or collaboration behavior. Hence, to interpret the results of FMC in respect of other effects due to changes in pure statistics like article, author or citation counts, we need a baseline to which we can compare the data. To acquire such a baseline in the following section we propose a modeling approach that takes into account the previously mentioned quantities.

7.3.2 Simulating article citations

We have seen that the fraction of matched citations varies for different PACS classes and different time windows. To be able to assess if the values of the FMC measure are significant or could be expected at random due to the given statistics of links and nodes we need a comparable baseline. This baseline or *null model* should consider all of the statistics and properties given by the data except the one we are focused on, which is the amount of matched citations. A common approach to such problems is an ensemble analysis that can be constraint by some desired parameters and allows to explore others

not included in the conditions.

However, to create the ensemble we first have to construct a feasible statistical model that describes the formation of links. The main effect on which we focus is the fraction of matched links. Therefore we fix all other properties of the network. In particular we keep the amount of in-coming and out-going citations for each article node. Further, we only allow citations between any two articles that preserve causality and therefore the cited article has to be published before the citing article. By fixing the in- and out-degree of all article nodes we also fix the amount of total citations in the network. The only freedom that remains in this model is exactly which nodes are connected to each other.

We denote the model that generates random realizations for given in- and out-degree sequences and preserving causality by $\mathcal{M}$. The pseudocode for this generative model is listed in Algorithm 1 and works as follows: We start with an empty network containing all article nodes that are present in a given time-window. First we choose a citing article node p_1 randomly with probability equal to its out-degree. Next we select a cited node p_2 with probability according to its in-degree from the subset of all articles that were published before p_1. We create the citation link (p_1, p_2) and lower the out-degree and in-degree of p_1 and p_2 by 1. We repeat this process until the out- and in-degree of all article nodes is equal to zero. During this process it can happen that we end up in a situation where we chose a citing article for which we can not find anymore a cited article that was published earlier. In this case we choose another cited article at random, thus violating the causality. However, we argue that only a few links will be affected and this will not alter the global topology significantly. In the end the algorithm returns a list of citations among the given articles nodes that preserve all of the initially fixed statistics.

Input : articles with their in- and out degree and publication date
Output: article citation links
begin
  Generate link set;
  $Cit \longleftarrow \emptyset$;
  Generate degree multi-sets;
  $Out \longleftarrow \emptyset$;
  $In \longleftarrow \emptyset$;
  for $a \in articles$ **do**
    for $i = 0$ **to** $deg_{out}(a)$ **do**
      Out.append(a)
    end
    for $j = 0$ **to** $deg_{in}(a)$ **do**
      In.append(a)
    end
  end
  while $Out \neq \emptyset$ **do**
    choose random $citing$ from Out;
    choose random $cited$ from In with $t(cited) < t(citing)$;
    add link $(citing, cited)$ to Cit;
    Out.remove($citing$);
    In.remove($cited$);
  end
  return Cit
end

Algorithm 1: $\mathcal{M}$: In- and out-degree preserving article citation network

The generation of articles takes all network statistics into account except exactly which articles cite each other. The citation in- and out-degree, the total amount of citations, the amount of articles and the causality of links remains unaltered. However, the particular pair of articles that is connected influences if a citation links is matched or unmatched and therefore effects the total fraction of matched citation FMC. We run the model $\mathcal{M}$ multiple times and compute the value of FMC for each realization. In Figure 7.4 we plot the average value of FMC as dashed lines over 100 simulations. A plot scaled to the simulation results solely can be found in Appendix B.2.

Strikingly, the average values of FMC for all the PACS classes are significantly smaller than the real FMC values. The simulated values of FMC reach at most 5% whereas the real values of FMC are all larger than 5%, except for PACS 80. We see that the

differences between the PACS classes are similar in the simulations and the real data. This is most likely due to topological features or network statistics. For example PACS 50 has the highest FMC values expected by the simulations and also the highest real FMC values. In cases of PACS 50 this can be explained by the comparably small amount of authors that are more densely connected to each other from a collaborative perspective. Due to the few authors there are also less articles published which makes it more likely for a article citation to be matched by a previous co-authorship. However, also in this case the real FMC is significantly larger than would be expected from random citations.

Since the model $\mathcal{M}$ preserves the degrees for each node individually the only reason for this can be the citation behavior. Such high amount of matched links could not be expected at random which implies that the social ties implied by previous co-authorships alters the probability of a citation in favor of a previous collaborator. Again, the absolute values of FMC vary for different PACS, but the average of random realizations is in all cases smaller than the real FMC values. From this we conclude that co-authorships can directly effect the amount of citations an article receives. This is in line with previous findings of the correlation of citation and co-authorships [98, 103, 143]. To analyze the citation behavior in scientific communities with comparably large FMC this social influence should be considered.

Further, the increasing FMC values can be interpreted as a signal of increased filtering and clustering of the scientific landscape. In terms of scientific knowledge transfer this means that there is a bias towards collaborators which makes it harder for scientist that have no or only few social ties to be recognized.

7.4 Multi-layer perspective on ranking schemes

Ranking schemes play an important role in estimating *scientific quality*. For hiring and funding committees it is important to have tools at hand that allow to identify outstanding scientist or research institutions. However, a generally accepted and objective measure of quality is still not available. Usual attempts in this direction either rely on bibliographic data, surveys or both of them. Even though very elaborate measures and indicators are developed, the weighting of different aspects is often crucial for the final ranks. There is no *ground truth* that allows to judge if a given ranking is meaningful or just some number crunching. Therefore simple indicators are often widely applied, taking the risk of inaccurate results, but avoiding the hassle to provide reasonable but complicated justifications.

As an example, rankings of universities gained a lot of attention in past years [62]. Especially, global rankings try to measure excellence of different research institutions. Among

the most famous ones are the Times Higher Education (THE) and the Academic Ranking of World Universities (ARWU). Not only gain these rankings a lot of interest from the public but also governments and policy makers keep track of the results [132, 161].

However, these rankings are often confronted with criticism in respect of their design and impact of different factors. The lack of precise methods and generalized statistics as well as the interpretations are criticized [162]. In the two aforementioned rankings also opinions and survey among staff and students are included together with bibliometric data. However the mixture of this factors and the weighting can lead to arbitrary results.

A different approach takes the Leiden ranking [170]. It focuses exclusively on bibliometric indicators of the publications affiliated with a university. This way they intend to emphasize research performance rather than reputation. Special attention is given to renormalization of citation data according to different research fields and publication types. Counting and normalization methods are another important concern that can alter ranking quite significantly [51, 171].

Also in the Leiden ranking and other quantitative approaches it is unclear what the actual implications of different ranking positions are [107]. It is suggested that confidence intervals for the uncertainty that underly different approaches should be considered instead of aggregated values [54]. To streamline the various research on theses topics 10 principles were established [63] that intend to assure correct use of research metrics and the accurate application to data.

Here, we critically discuss ranking schemes based on bibliometric data that use the pure amount of citations as the main indicator. We use the multi-layer perspective to point out the influence of co-authorships. Further, we investigate how the aggregation of article data to the layer of authors and affiliations leads to a loss of the article topology. To analyze this relationship we introduce a generative model that incorporates the information available from the multi-layer perspective.

In the previous section we discussed how co-authorships can influence later citations between collaborators. We found that a significant fraction of citations preferably happen between articles published by authors which collaborated before on a common article. Notably, this can not be expected at random even when correcting for the in- and out-degrees of all articles. Leveraging on this findings in a next step we intend to incorporate the social influence to develop a model for generating article citations. This model should take into account all effects that we can measure and that are not necessarily bound to scientific quality of an article. Using such a model as baseline gives us the opportunity to identify outliers that significantly deviate from the model prediction.

7.4.1 Article citation model

An appropriate model to generate article citations should preserve the measurable statistics of the data and allow freedom for the properties one wants to investigate. The main focus of our generative model is the amount of incoming citations of an article. We want to know how many citations a certain article can expect to receive at random considering some boundary conditions. These boundary conditions should include all statistics of the citation network that can be measured based on the data. These statistics are usually not related to scientific quality and therefore should already be included as a factor that influences the formation of citations.

One of the ingredients of the model should be the publication date of an article. This accounts for the fact that older publication are more likely to be cited by the plenty amount of new publications. Further we preserve the amount of references of each article. This means that the out-degrees of all citations remains unaltered. However, to which articles this citations go are up to the model randomization. Since, we learned from the previous section that co-authorships influence citation behavior we also include this bias in our model. More precisely, for each article in the data we first identify how many out-going citations are matched and how many are unmatched. Hence, we also preserve the absolute values of matched and unmatched out-going citations for each article individually. Therefore this model not only reproduces the fraction of matched citations in the whole article citation network, but also the citation composition of each single article. This means, that each article in the model will cite the same amount of articles as in the data and also the fraction of matched citations will be the same.

We denote the stochastic model generating an article network in the previously described way by $\mathcal{CM}$. The pseudocode is listed in Algorithm 2 and goes as follows: First we generate the matched article network $G_{matched}$ that contains all matched citations that are possible. For each article node a we observe the amount of matched citations in the data. Then we randomly choose the same amount of articles from the neighbors that a has in $G_{matched}$, thus assuring that we only choose cited articles that lead to matched citations. Next we compute the difference of the out-degree of a and the amount of matched citations of a. This number gives us the amount or remaining articles to be chosen such that no matched citation is created. In the end the algorithm returns citations among the initial set of article nodes that preserve the desired statistics.

Input : articles with out-degree, matched citations and publication date
Output: article citation links
begin
  $Cit \longleftarrow \emptyset$;
  $G_{matched} \longleftarrow$ matched article network;
  for $a \in articles$ **do**
    for $i = 0$ **to** $MatchedCount(a)$ **do**
      choose random neighbor b of a in $G_{matched}$;
      add link (a, b) to Cit
    end
    for $j = 0$ **to** $(deg_{out}(a) - MatchedCount(a))$ **do**
      choose random article b that is not a neighbor of a in $G_{matched}$;
      add link (a, b) to Cit
    end
  end
  return Cit
end

Algorithm 2: Stochastic article citation model $\mathcal{CM}$ that preserves the out-degree and fraction of outgoing matched citation for each article.

All networks that are generated by $\mathcal{CM}$ preserve the citation statistics extracted from the data. More precisely, this comprises the out-degrees, matched citation count per node and publication date. The stochastic model $\mathcal{CM}$ gives rise to an ensemble. By running the model multiple times we can empirically compute the average citation count an article receives at random under the constraints of $\mathcal{CM}$. We denote the average citation in-degree of an article node a by $\langle cit_{in}^{\mathcal{CM}}(a) \rangle$.

The generative model $\mathcal{CM}$ allows to estimated the amount of citations an articles receives at random also considering the social bias. The difference of the actual citations an articles receives to the expected citation based on $\mathcal{CM}$ allows to measure outstanding articles. This leads to a measure of quality that is adjusted for the measurable statistics of links and inter-layer correlations we acquired from the multi-layer perspective. Based on the generative model $\mathcal{CM}$ we will discuss citation based rankings in the following section.

To evaluate if the inclusion of matched citations is relevant we compare the results of $\mathcal{CM}$ to a so-called *null model* $\mathcal{CM}_0$. The null model works in the same way as $\mathcal{CM}$ with the only difference that it does not distinguish matched from unmatched citations and therefore does not include any social bias implied by co-authorships.

7.4.2 Ranking schemes

Considering the article citation model described in the previous section we have a method available to estimate the citations of articles. Most rankings nowadays are still based on the amount of citations a scientist or university receives. Even though other indicators are included as well, citations still play an important role. Hence, we compare the citations based on the generative model $\mathcal{CM}$ to the actual citations in the data. We apply the commonly used Pearson and Kendall-Tau rank correlation measures to identify how well the estimations of $\mathcal{CM}$ coincides with the actual data. While the Parson measure gives the linear correlation of the values themselves, the Kendall-Tau measure focus on the ranking of the values rather than the actual size. Both measure range between -1 implying anticorrealtion, to 1 implying complete agreement, whereas 0 implies that the values are uncorrelated.

Our analysis is based on each PACS class individually. We focus on the last time-window in our data that spans from 2005 to 2010. Note that we still do not consider self-citations and therefore neglect all citations where there is an overlap of authors of the citing and cited article. The simulations are done on the article layer, however we project the article citation links to the author and affiliations layer. This means that for the author layer we evaluate the amount of author citations they receive based on the data or the generative model $\mathcal{CM}$. The same is done for the affiliation layer, where we evaluate the citations between affiliations. We apply a fractional counting to the projected citation links to account for the fact that one article citation link results in several author and affiliation citation links. The difference between the layers is due to an aggregation of the article layer topology. By this we mean that an author can publish several articles and all the citations these articles receive are aggregated to one author. For affiliations the aggregation is even stronger since all articles of all authors of an affiliation are aggregated to one affiliation node. The results are listed in Table 7.1. The values are given separately for each PACS class and each layer, articles, authors or affiliations. Finally, we also include the results for the null-model $\mathcal{CM}_0$ to identify the sole impact of the co-authorship bias.

First, we can observe that there are differences between the various PACS classes but that the values are nevertheless comparable. Focusing on the article layer we observe that the Pearson correlation ranges from 0.22 up to 0.33 and the Kendall-Tau correlation ranges from 0.30 up to 0.36. This values are not that high, implying that in the real citation network there are biases included that are not captured by the $\mathcal{CM}$ model. We would also expect this, since the generative model only considers the publishing date, the total citation count and the social bias. Everything else such as scientific quality or impact of the articles is not covered by our approach. However, a minor but significant part

PACS:	**00**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**
Model $\mathcal{CM}$:										
Articles										
Pearson	0.27	0.29	0.30	0.30	0.26	0.33	0.26	0.25	0.22	0.31
Kendall-Tau	0.36	0.33	0.34	0.32	0.36	0.30	0.31	0.36	0.33	0.33
Authors										
Pearson	0.68	0.64	0.70	0.64	0.66	0.63	0.61	0.73	0.58	0.67
Kendall-Tau	0.53	0.55	0.50	0.47	0.49	0.42	0.42	0.53	0.43	0.53
Affiliations										
Pearson	0.92	0.92	0.95	0.92	0.92	0.90	0.95	0.95	0.88	0.89
Kendall-Tau	0.75	0.72	0.69	0.69	0.72	0.59	0.72	0.77	0.69	0.70
Null Model $\mathcal{CM}_0$:										
Articles										
Pearson	0.11	0.12	0.13	0.13	0.11	0.13	0.12	0.11	0.10	0.13
Kendall-Tau	0.17	0.15	0.16	0.17	0.17	0.14	0.15	0.17	0.16	0.16
Authors										
Pearson	0.64	0.56	0.64	0.61	0.62	0.54	0.59	0.70	0.55	0.65
Kendall-Tau	0.52	0.52	0.48	0.47	0.48	0.38	0.41	0.52	0.42	0.51
Affiliations										
Pearson	0.91	0.90	0.95	0.91	0.91	0.87	0.95	0.94	0.87	0.88
Kendall-Tau	0.74	0.71	0.67	0.68	0.71	0.57	0.72	0.77	0.69	0.69

Table 7.1: Correlation results Pearson and Kendall-Tau rank correctional of citation in-degrees of the data compared to simulations based on the model $\mathcal{CM}$ and the null model $\mathcal{CM}_0$. The correlations are shown for the article, author and affiliation layer. The evaluation is based on articles published between 2005 and 2010. The values represent averages over 100 simulations.

of the citations could still be expected at random and therefore are not relevant for the identification of outstanding articles.

If we compare the correlations values of the articles layer based on $\mathcal{CM}$ to the null model $\mathcal{CM}_0$ we see that the social bias implied by co-authorships is significant. For the null model the Pearson correlation ranges between 0.10 up to 0.13 and the Kendall-Tau correlation ranges from 0.14 up to 0.17. Hence, the agreement of the null model with the actual data is much worse than the agreement of $\mathcal{CM}$ with the data. Since the only difference of the two generative models is the inclusion of the co-authorship bias this implies that matched citations are indeed relevant for the formation of article citations.

Let us look now at the differences of the correlation values of the article layer to the author and affiliation layer. We can see that the values significantly increase for the author layer and become larger for the affiliation layer, for all of the PACS classes. Recall that the

author and affiliation layers are both constructed based on the article layer an there is no additional information included considering the citation statistic. The only difference comes from the information which authors were involved in which articles. From the perspective of random citations, authors that published a lot of articles are more likely to receive citations. The same happens for affiliations to which a lot of articles can be attributed. Hence, the mere amount of articles that correspond to one affiliation increase the probability of citations that could be expected at random. This is a so-called *size effect*. Based on the construction it can be expected that the size effect is present in the generative model. However, the higher correlation values, especially for the affiliations, indicate that the size effect is also very prominent in the real world citation network. The size effect almost makes the social effect induced by co-authorships negligible. Nevertheless, the correlation values of $\mathcal{CM}$ are still equal or higher than the correlation values of $\mathcal{CM}_0$ even for the author and affiliation layer.

Summarizing, we can conclude that the citation count of authors or affiliations is heavenly influenced by the size in terms of articles attributed to them. This means that if we measure the amount of citations an affiliation receives we are merely measuring the amount of articles that are published by this affiliation. However, the amount of articles is not necessarily an indication for scientific quality. Therefore one should in general be very careful when citation counts of affiliations are used as ranking criterion.

7.4.3 Surprise factor

Even though the citation bias implied by co-authorships is mitigated on the author and affiliation layer it is significant for the article layer. Hence, for each article node p we can calculate the difference of the citation in-degree of the simulation results $\left\langle cit_{in}^{\mathcal{CM}}(p)\right\rangle$ to the actual citation in-degree $cit_{in}(p)$ of the data. We call this difference *surprise factor* denoted by $\mathcal{SF}(p) = cit_{in}(p) - \left\langle cit_{in}^{\mathcal{CM}}(p)\right\rangle$. Since the stochastic model $\mathcal{CM}$ actually results in a distribution of different citation in-degrees values for each article, we can also consider the significance of the simulation results in respect to the data. A measure that can be used in such cases is the z-score, $z = \frac{x-\mu}{\sigma}$, that evaluates the difference of the true value x to the sample mean μ in respect of the standard deviation σ. The z-score is based on normal distributed samples, however, it can also be adjusted to other distributions. In our case the citation for individual nodes are not exactly normal distributed but similar. Since we only want to provide an idea how the surprise factor can be applied we do not give a technical justification. Therefore, an adjusted surprise factor that also accounts for

the variability of the simulations could be defined as follows,

$$\mathcal{SF}(p) = \frac{cit_{in}(p) - \left\langle cit_{in}^{\mathcal{CM}}(p) \right\rangle}{\mathrm{Std}(cit_{in}^{\mathcal{CM}}(p))} \,. \tag{7.4}$$

Independent of which definition we use for the surprise factor, it tells us how *unexpected* the real citation count of an article is compared to the generative model $\mathcal{CM}$. A positive $\mathcal{SF}$ value for an article indicates that it received more citations than we would expected at random and therefore is possibly an outstanding article. On the other side if $\mathcal{SF}$ value is negative for an article it indicates that it received less cations that one would expect at random.

Hence, the surprise factor could serve as indicator that highlights articles that deviate in terms of citations from what would be expected at random considering the co-authorship bias. Of course the surprise factor should not be applied as standalone measure but could be considered as an additional source of information in combination with other indicators. Especially, it is not suitable to build a ranking just based on the score of $\mathcal{SF}(\cdot)$. However, in case of dyadic comparison it could be insightful. Consider two articles p_1 and p_2 with the same citation counts, i.e. $cit_{in}(p_1) = cit_{in}(p_2)$, but with different surprise factors $\mathcal{SF}(p_1) > \mathcal{SF}(p_2)$. This would imply that articles p_1 received more unexpected citations than p_2. If both articles were published on the same date this is due to the fact that p_2 received more matched citations, meaning that more citations came from previous collaborators. This information could be helpful in evaluating if an article was qualitative enough to also receive acknowledgment from scientists which had no previous contact with the authors of the article.

As with most indicators one has to be careful with the interpretations and be aware what exactly the surprise factor captures. One should especially be cautious since scientists that maintain a huge network of collaborators could have a disadvantage from this measure. Therefore, we do not claim that a high surprise factor is either good or bad, but just offers an additional perspective on the relation of co-authorships and citations.

7.5 Conclusion

In this chapter we have investigated the social influence of collaborations between scientists on their citations. We hypothesized that due to the vast amount of scientific publications that become available, scientists can not track anymore all of the relevant works. Hence, they rely on some filtering mechanism to simplify the selection process. We consider the social contacts to other scientist as such a filter strategy. To test our assumption

we analyzed how collaborations, which indicate social contacts, influence citations, which represent the acknowledgement of other works. Our study is based on articles published in APS journals.

More precisely, using a multi-layer approach we analyzed how co-authorships influence the occurrence of citations between authors. We distinguished two types of relations, co-authorship and citation links. Using correspondence links that connect articles to their authors and affiliations, we were able to project co-authorships that happen between authors to the layer of articles and affiliations. In this way we identified matched citations between articles that imply that at least one author of the citing article previously co-authored an article with at least one author of the cited article. We introduced a measure denoted by FMC that quantifies the fraction of matched citations in a topical constrained subfield of articles. Hence, FMC indicates how many citations are possibly due to social contacts that some authors of the citing and cited article maintained. To evaluate if FMC can be interpreted as a measure for a social bias or could be expected at random, we compared it to a random generative model. This stochastic model took all bibliometric statistics of the data into account to rule out any random effects which could be expected. However, we found that the amount of matched citations was significantly higher than could be expected at random in all topical fields that we investigated.

This finding verifies our initial hypothesis and shows that there is a social bias that influences citations. Furthermore, the value of FMC increased over the past two decades which implies that this bias becomes more and more pronounced. Potentially, this leads to a filtering of the scientific awareness, meaning that acknowledgments are tailored towards the works done by scientists that know each other. To quantify this effect we provided a simple measure that allows to further study how knowledge transfer is affected.

Considering the social influence of co-authorship onto citations we discussed citation based ranking schemes. We suggested a generative model for article citations that consider the impact of co-authorships. Using the average citations generated by this model as a baseline we evaluated the agreement with the real data. We found that to some extend the model is able to explain the real amount of citations an article receives. Nevertheless, there is still a lot of topological structure in the citation network that can not be explained by random citations. Hence, we used the deviation from the data to the model expectations to define a surprise factor that serves as estimation of the quality of an article.

However, we found that projections of citations of articles to their corresponding authors or affiliations lead to much higher correlations with the data. This is due to a size effect, which means that a single author or affiliation can be connected to several articles, which in return increase the probability to receive a citation at random. In other words this

means that citations merely capture the number of publication associated with authors or affiliations. While it is feasible that the size effect is present in the generative model, the agreement with the real data implies that the mere amount of articles of an author or affiliation is also the driving factor of real citation counts. Especially for affiliations we conclude that a ranking based on citations is not feasible to make statements of quality.

However, for the citations of articles the surprise factor can be used to assess the influence of co-authorships. In particular we can distinguish two articles that exhibit the same citation count but different surprise factors. Meaning, that we can evaluate if citations are matched by previous collaborations or come from socially unrelated authors. The surprise factor should not be used on its own but provides an additional perspective on citation based indicators and ranking schemes.

Concluding, in this chapter we provided several indicators to measure the social influence of co-authorship onto citations. Our analysis was based on a multi-layer perspective that allows to combine multi-dimensional relations. Only observing each layer as a separate network would not have allowed to reveal the hidden correlations that are present in the data. The methodology presented in this chapter can also be applied to other aspects of scientometric investigations. Considering interrelated aspects is in general a promising approach and provides new quantification tools for the science of science.

Chapter 8

Conclusions

8.1 Summary and Discussions

The goal of this thesis was to explore and develop new methods in network theory that go beyond the standard approach to study complex systems. In particular we intended to build an extended framework that overcomes some of the limitations imposed by basic static one-dimensional network representations. More precisely, we investigated the inclusion of time with temporal networks and analyzed interconnected and multi-dimensional systems with multi-layer networks. We focused on specific aspects of the aforementioned topics to provide further insights and suggest new approaches to some open issues. We provided novel higher-order network models that capture properties of complex systems that are neglect by standard network approaches.

8.1.1 Part I: Temporality

In Part I we focused on temporal ordering of links in temporal networks. Temporal networks include a time-stamp for each link therefore indicating at which points in time a link was present. In this framework we studied the order in which interactions occurred or links were traversed. The ordering can have a crucial impact on the topology and path structures of temporal networks. It can be based on time-stamped links or path statistics gathered from the data. Casual methods usually aggregate time-stamped links for given time windows thus neglecting any kind of time dependency. Such networks are called first-order aggregate networks or static networks. To overcome this simplification in Chapter 3 we first introduced a framework to capture path statistics that explicitly consider the order of links. The so called k-order aggregate networks capture the correct occurrences of paths

of length k in the temporal network. These higher-order networks are also aggregations of the real temporal sequence but with the addition that path statistics get captured as well. To highlight the importance of a correct representation of time-dependent paths we focused on two applications of our framework.

In Chapter 4 we investigated path-based centrality measures. Since time-respecting paths may not get captured accurately by a time-aggregated perspective this also influences the position of nodes in respect to their neighbors. Centrality measures on static networks implicitly assume that all paths that exist in the static time-aggregated network also exist in the temporal sequence. Since this is not true in general we first defined temporal versions of the commonly known betweenness, closeness and reach centrality. The temporal centralities explicitly preserve the link ordering of all paths in the temporal sequence. Next, we defined centrality measures based on the higher-order aggregated networks. We compared the centrality measured based on second-order and first-order aggregate networks to the temporal version. We found that higher-order aggregate networks better capture the true temporal centralities than assumed by static time-independent representations.

In Chapter 5 we focused on the impact of ordering on dynamical processes running on temporal networks. We highlighted the relation of the link ordering to temporal causality and how the non-Markovian property of consecutive links in temporal sequences compare to a first-order time-aggregated perspective. This means that in the empirical temporal sequence the occurrence of a link $b \to c$ with the source b does also depend on a previous links $a \to b$ with target b. Such dependencies between sequences of links influence how one can navigate in a temporal network and therefore how something spreads through the network. In particular we focused on a random walk process that models diffusion dynamics and can be used as a proxy for several spreading processes in networks. In terms of temporal causality we focused on second-order networks that capture paths of length two and therefore the dependency of links on one previous link. This can also be considered as one-step memory of a random walker. To estimate the amount of non-Markovian two-paths in the temporal sequence we used an entropy measure that indicates how much freedom a walker has in choosing his next step. We empirically investigated how fast a random walk converges on a second-order aggregate network compared to a first-order aggregate network. We found that the inclusion of time-respecting paths can either slow-down or speed-up the random walk process compared to a static time-aggregated network. Further, we provided an analytical measure, called slow-down factor, which accurately predicts the change in diffusion speed from a second-order to a first-order perspective. We finally analyzed some topological properties such as community and geodesic structures in the network which allow to explain some aspects that alter the diffusion speed in a temporal network compared to a static representation. We also briefly described methods

to reorder time-stamped links in a temporal sequence to either slow-down or speed-up a diffusion process in a temporal network.

Conclusively, in Part I we showed that the methodology of higher-order aggregate network is a powerful tool to analyze the ordering of links in temporal networks. Note that there are some crucial conditions that have to be considered in our framework. First of all we only consider temporal networks composed of discrete time-steps. Hence, the results derived here may not directly translate to continuous processes. However, temporal networks based on real data are usually also gathered from discrete observations which makes our approach a natural application to real world data sets.

Another crucial aspect in the construction of higher-order aggregate networks is the choice of an appropriate and suitable maximal time difference δ that defines the time-span in which consecutive paths are considered. While a too large δ aggregates the time-stamped links therefore losing some of the order correlations, a too small δ renders the temporal network disconnected and only allows few paths to exist. In our analysis we argued for an intermediate δ related to a strongly-connected network perspective where all nodes have the possibility to influence each other by a time-respecting path. In general an appropriate δ should be chosen according to temporal characteristics of the underlying real world system, that is represented by a temporal network, and the time scale of the dynamical process. However, the exact choice of a maximum time difference δ may be a difficult and ambiguous task in particular cases.

The construction of higher-order aggregate networks and their analytical investigation may be computationally expensive for large temporal networks. The amount of nodes in k-order aggregate network scales with the amount of paths of length k in the temporal network. For example, the size of a second-order time-aggregated network scales with the amount of links and the amount of second-order links with the amount of two-paths. However, we argue that real world temporal data is often sparse and therefore not all possible paths have to be considered. Further, the use of higher-order aggregate network is still more efficient than considering the whole temporal sequence of links and therefore is a viable proxy for the real temporal order correlations.

We conclude that our analysis of temporality of complex systems provided new insight that go beyond the standard network perspective. Our framework to study the ordering of temporal links provided a novel approach that includes causality and correlations effects that were widely neglected before. Finally, we provide measures that allow to identify and quantify the effect of temporal ordering in any given data set, thus providing valuable methods for a lot of applications.

8.1.2 Part II: Interconnectivity

In Part II we focused on interconnected and multi-dimensional systems. We used the multi-layer framework to represent networks with certain topological properties that are connected to each other, and interconnected networks that combine multiple node and link types. Since the possible applications of multi-layer networks are very broad we focused on particular subjects that highlight the aforementioned aspects.

In Chapter 6 we theoretically investigated the lack of knowledge in interconnected networks. If the topology of the whole network is available there would be no difference to a single layer network that is composed of several components. However, here we assumed situations where we lack knowledge of either the intra-link topology of single layers or the inter-link topology between layers. These situations are motivated by real world systems, for example by online social networks, where the exact knowledge on how different systems are connected is not available and one has to rely on some aggregate statistics. To deal with this lack of knowledge we used an ensemble approach to estimate average properties of several realizations of a multi-layer networks that either coincide with the inter- or intra-link topology. Our approach had the purpose to estimate the properties of a dynamical process. In particular we focused on a random walk process that runs on a multi-layer network. We analyzed how the convergence speed of this random walk process relates to either knowing the intra- or inter-link topology of the multi-layer network. By varying the density of intra- and inter-links we identified cases where the aggregate statistics are enough to precisely estimate the convergence of a random walker. In some cases it is even enough to consider a multi-layer aggregation that only considers the densities of both, inter- and intra links.

In this chapter we mainly derived theoretical results for randomly generated networks. Layers that contain Erdös-Rényi and regular networks exhibit link topologies that are well suited to differentiate the situations where the ensemble predictions work best. Random walks on network layers the exhibit scale free link topologies need usually quite long to converge to a stationary state resulting in relatively small difference between the approaches. Nevertheless, the methodology works in general and could be applied to any kind of network as as long we have a suitable stochastic model underlying it.

In Chapter 7 we analyzed a real multi-relational data set of scientific publications. We intended to investigate the correlation of citations and co-authorships between scientists. We hypothesized that due to the increasing amount of publications over the last decades, scientist more often rely on the social contacts to filter the relevant information. To test this hypothesis we interpreted co-authorships as social contacts of collaborations and ci-

tations as acknowledgement of relevant work. The relation of articles to their authors and affiliations allowed to build a multi-layer network interconnected by their correspondence. The intra-links represent either citations or co-authorships. We developed a random citation model that allowed to find the average citation an article receives at random while preserving publication dates as well as in- and out degrees of article nodes. We defined the term of matched citations indicating that at least one author of the citing article previously collaborated with at least one author of the cited article. Analyzing subfields based on APS we found that the fraction of matched citations is much higher than would be expected at random, thus verifying our assumption. Knowing that co-authorships can alter citation behavior we constructed an article citation model to estimate the amount of citations an article receives incorporating the collaboration topology of authors. Our model had a good correlation with the actual citations but still allowed us to identify articles respectively authors and affiliations that perform better than expected. Building on this we provided an new perspective on ranking schemes that take into account the correlation between citation and co-authorships.

In both parts we established new methods to tackle various issues in the analysis of complex systems. The findings are therefore relevant for various scientific fields.

8.2 Scientific contribution

Her we briefly discuss how our findings are of relevance to particular research fields. All of our results are contributions to network science and the general analysis of complex systems. The framework of higher-order aggregate networks, to deal with causality in time-stamped data, is a general approach that can be applied to any kind of ordered data. This also applies to the centrality measure and slow-down factor that we developed in this respect. The results on the diffusion of interconnected networks are theoretical and therefore applicable to any kind of interconnected network where the link type is the same across and within the network. The methodology used to analyze data on scientific citations and collaborations were tailored for this particular investigation, however, the idea of constructing an ensemble that considers a multi-layered structure can also be of use to other applications that include multi-dimensional relations.

So far, three publications cover some parts of the research presented in this thesis and make the results available to the scientific community. The methodology of higher-order aggregate networks was published in *Nature Communications* [144] and *EPJ-B* [145]. *Nature Communications* is an interdisciplinary journal that publishes research from all areas of the natural sciences that represent important advances of significance within specific dis-

ciplines. *EPJ-B* is part of the *European Physical Journal* specialized on condensed matter and complex systems. Both these publications highlight the importance of the presented higher-order approach for various scientific fields that go beyond network science. The ensemble approach to study diffusion in interconnected networks was included in a chapter [175] of a joint book on multi-layer networks called *Interconnected Networks* [48]. This book is part of the dissemination of the European project *MULTIPLEX* dealing with the various topics of multi-level complex systems. The research on the correlation citations and co-authorships has yet to be published and is planned to target the research field of scientometrics.

More specific contributions are briefly outlined in the following:

Data Mining Data mining is a branch of computer science that deals with processing and analyzing large data sets. This field is constantly interested in efficient representations and tools to analyze huge amount of data. In terms of analyzing time-stamped data, higher-order aggregate network provide a balanced compromise. They capture the causality of consecutive links but still do not require to keep track of the detailed temporal sequence. It also allows for stochastic investigation and delivers path statistics that are in line with the real data. For larger networks the computational cost may increase but are still far better than considering the detailed temporal data. Also taking into account the slow-down factor and entropy growth rate ration, our framework provides an efficient and analytical substantiated approach to analyze large amount of temporal data.

Biological Pathways The network perspective has proved to be useful for the analysis of drug actions and disease complexity as well as drug design [30]. How molecules interact with each other or how genes are activated is crucial in understanding disease and providing appropriate treatment. Correct ordering and causality is of special interest for biological pathways. Exact pathways of biological interaction are essential to understand metabolism and the regulation of gene expressions. Due to the complex structure common approaches still aggregate the biological pathways over time to receive statistics of molecule and gene activations but losing the exact ordering. Therefore, a higher-order aggregate network approach could deliver a more detailed analysis that considers the exact sequences of activations an therefore allows to represent real biological pathways more accurately.

Ranking Schemes The ranking method presented in Chapter 7 provides a alternative perspective on citations based rankings. Our method can be applied to ranking articles, scientist or institutions. Especially, university ranking got a lot of attention in recent

years [62]. They consider a lot of different indicators to measure excellence of researchers. Even if they consider citations and co-authorships in their ranking schemes they often treat them as separate parameters. However, the assumption of independence is not feasible in general. Further, we have seen that due to the size effect citations based ranking have to be carefully evaluated on higher aggregation levels. To partially overcome this issue, our approach highlights that investigation and models of citations should be based on links between articles. Aggregations to scientist or institutions should only be done as projection of the result acquired from the articles. To rule out other statistical properties, which should not be considered for scientific quality, the surprise factor offers an additional indicator to identify outperforming works. Summarizing, our methods provide an alternative perspective on rankings and allow to incorporate correlations with indicators based on stochastic modeling.

8.3 Outlook

In this thesis we investigated methods and models based on network science to target issues in the analysis of complex systems. Due to the vastness of this field we could only focus on few particular topics. We provided new insights and some general methodology that can be used in various applications. However, there are several potential extensions to the methods and findings discussed in respect to temporal and multi-layer networks.

In regard of temporal networks the right choice of a maximum time difference δ is still an open issue. Recall that δ is the maximal amount of time that can pass between two links until they are not considered to be part of a time-respecting path anymore. The right choice of such a time window is not only of relevance of the higher-order aggregate network presented in this thesis but also to the analysis of dynamical processes in general. The particular choice of δ in our research was based on strongly-connected networks to allow a feasible analysis of time-respecting paths. However, future research should investigate in more detail how an appropriate δ should be chosen depending on the dynamics and system of interest.

In regard of temporal centrality measures there are several opportunities for expansion and development of new methods. First of all, in our investigation we focused on three path-based centrality measures that did not consider link weights. However, the inclusion of links weights and the expansion to other centrality measure could lead to promising results. More complex measures like temporal eigenvector centralities or temporal PageRank [118] could provide more accurate representation of navigation patterns in temporal networks.

The results on temporal dynamics and the presented slow-down factor where based on a

random walk process. However, the causality of temporal ordering is also of significance to other dynamical processes. In general it is still a popular opinion that temporal dynamics slow down a process, therefore further investigation should focus on potential speed-up occurring for other dynamical processes. Additionally, the methodology of temporal re-ordering of links to slow-down or speed-up a process could be applied to a real-world system that can be designed in an appropriate way.

Even thou we presented the framework of k-order aggregate networks we merely focused on the inclusion a second-order links. For some application this makes scenes since it can be regraded as one-step memory, that is somehow natural in social interaction processes. However, it is up to further investigations how the consideration of longer path lengths and combination of them influence or improve our results.

The ensemble approach to study the lack of knowledge in multi-layer networks was based on a theoretical analysis and simulation results. Nevertheless, we foresee potential of an actual application to real-world scenarios that exhibit the particular topologies considered by our framework. Adjustments of the ensemble approach may further allow to study other processes that are based on an analytical framework.

Our results and elaborations on the multi-layer perspective of citation and co-authorship between scientist call for more detailed research. The multi-layered perspective allows to disentangle the correlations of various citing behaviors. Even though we discussed how rankings are influenced by mere citations and co-authorships we did not provide a conclusive answer how one should deal with this. Therefore, further research for multi-layer ranking mechanism could offer a promising extension of classic oversimplified ranking schemes.

Finally, the development of novel higher-order network models to study complex systems is a never ending process. The increasing availability of large and more detailed data sets creates new challenges, which require an extend framework. Therefore, this thesis could only provided some starting points for further explorations of the network approach.

Appendices

Appendix A

Temporality

A.1 Derivation of slow-down factor

In chapter 5, we argue that changes of diffusion dynamics in temporal networks as compared to their static counterparts, are due to the change of *connectedness*, or *conductance*, of the corresponding *second-order aggregate network*. We further show that these changes are captured by a slow-down factor which can be computed based on the second-order aggregate networks corresponding to a particular non-Markovian temporal network and its Markovian counterpart. In the following, we substantiate our approach by analytical arguments, highlighting the conditions under which our prediction is accurate.

For a second-order aggregate network $G^{(2)}$ with a weight function $w^{(2)}$, let us consider a transition matrix $\mathbf{T}^{(2)}$ as defined in Eq. 2 of our article. The influence of the eigenvalues of $\mathbf{T}^{(2)}$ on the convergence behavior of a random walk can then be studied as follows. For a sequence of eigenvalues $1 = \lambda_1 \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$ of $\mathbf{T}^{(2)}$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, we define the *eigenmatrix* $\mathbf{U} := (\mathbf{v}_i)_{i=1,\ldots,n}$. We further define a stochastic row vector $\mathbf{x} = \boldsymbol{\pi}_0 = (p_1, \ldots, p_n)$ which we assume contains the initial node visitation probabilities before the random walk starts. Since $\mathbf{U}$ is not necessarily regular (n.b. that $G^{(2)}$ is directed) we use a Moore-Penrose pseudoinverse [124] $\mathbf{U}^{-1}$ of $\mathbf{U}$ as well as diagonal matrix $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ to obtain an eigendecomposition of $\mathbf{T}^{(2)}$ as

$$\mathbf{T}^{(2)} = \mathbf{U}^{-1}\mathbf{D}\mathbf{U}. \tag{A.1}$$

We can then transform the vector $\mathbf{x}$ into an eigenspace representation of $\mathbf{T}^{(2)}$ and obtain $\mathbf{a} = \mathbf{x}\mathbf{U}^{-1}$ such that $\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{v}_i$. With this, the node visitation probability vector $\boldsymbol{\pi}_k$

after k steps can be expressed as

$$\boldsymbol{\pi}_k = \mathbf{x}\mathbf{T}^k = \sum_{i=1}^{n} a_i \mathbf{v}_i \mathbf{T}^k$$

where $\mathbf{T}^k$ is the k-th power of the transition matrix $\mathbf{T}$ and a_i is the i-th entry of vector $\mathbf{a}$. Repeated substitution according to the eigenvalue equation $\mathbf{v}_i\mathbf{T} = \lambda_i \mathbf{v}_i$ yields

$$\boldsymbol{\pi}_k = \sum_{i=1}^{n} \lambda_i^k a_i \mathbf{v}_i.$$

Assuming that $\mathbf{T}^{(2)}$ is primitive, for the Perron-Frobenius eigenvalue λ_1 we obtain $1 = \lambda_1 > |\lambda_2|$ and the normalised first eigenvector $a_1\mathbf{v}_1$ corresponds to the unique stationary distribution $\boldsymbol{\pi} = \boldsymbol{\pi}_k$ of the Markov chain given by $\mathbf{T}^{(2)}$. For the first term in the sum above, we thus obtain $\lambda_1^k a_1 \mathbf{v}_1 = 1 \cdot \boldsymbol{\pi} = \boldsymbol{\pi}$. With

$$\boldsymbol{\pi}_k = \boldsymbol{\pi} + \sum_{i=2}^{n} \lambda_i^k a_i \mathbf{v}_i \tag{A.2}$$

a difference vector $\boldsymbol{\delta}(k)$ whose components $\delta_j(k)$ capture the difference between node visitation probabilities $(\boldsymbol{\pi}_k)_j$ after k steps of the random walk and the stationary visitation probability $(\boldsymbol{\pi})_j$ for each node j can be defined as

$$\boldsymbol{\delta}(k) = \boldsymbol{\pi}_k - \boldsymbol{\pi} = \sum_{i=2}^{n} \lambda_i^k a_i \mathbf{v}_i. \tag{A.3}$$

The total variation distance

$$\Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) := \frac{1}{2} \sum_{j=1}^{n} |\, (\boldsymbol{\pi})_j - (\boldsymbol{\pi}_k)_j \,|$$

after k steps can then be given as

$$\begin{aligned} \Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) = & \quad \frac{1}{2} \sum_{j=1}^{n} |\delta_j(k)| \\ = & \quad \frac{1}{2} \sum_{j=1}^{n} |\lambda_2^k a_2 (\mathbf{v}_2)_j + \lambda_3^k a_3 (\mathbf{v}_3)_j \\ & \qquad + \ldots + \lambda_n^k a_n (\mathbf{v}_n)_j \,| \end{aligned}$$

where $(\mathbf{v}_i)_j$ denotes the j-th component of the i-th eigenvector $\mathbf{v}_i$. Under the condition that $|\lambda_2|$ is not degenerate (i.e. $|\lambda_2| > |\lambda_3|$) and using the fact that $|\lambda_i| < 1$ for $i \geq 2$ (n.b. that $\mathbf{T}^{(2)}$ is primitive and thus $G^{(2)}$ is necessarily strongly connected) for k sufficiently large one can make the following approximation:

$$\Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) \approx \frac{1}{2} \sum_{j=1}^{n} |\lambda_2^k a_2 \left(\mathbf{v}_2\right)_j|.$$

For a sufficiently small convergence threshold $\varepsilon > 0$, the convergence time k after which the total variation distance falls below ε can then be calculated as follows:

$$\begin{aligned} \Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) \approx \frac{1}{2} \sum_{j=1}^{n} |\lambda_2^k a_2 \left(\mathbf{v}_2\right)_j| \leq \varepsilon \Leftrightarrow \\ k \cdot \ln(|\lambda_2|) + \ln\left(\frac{1}{2} \sum_{j=1}^{n} |a_2 \left(\mathbf{v}_2\right)_j|\right) \leq \ln(\varepsilon) \Leftrightarrow \\ k \geq \frac{1}{\ln(|\lambda_2|)} \cdot \left(\ln(\varepsilon) - \ln\left(\frac{1}{2} \sum_{j=1}^{n} |a_2 \left(\mathbf{v}_2\right)_j|\right)\right) \end{aligned}$$

Here, we utilise the fact that, since $|\lambda_2| > |\lambda_3|$, both λ_2 and $a_2\mathbf{v}_2$ are necessarily real and thus $|\lambda_2^k a_2 \left(\mathbf{v}_2\right)_j| = |\lambda_2^k| \cdot |a_2 \left(\mathbf{v}_2\right)_j| = |\lambda_2|^k \cdot |a_2 \left(\mathbf{v}_2\right)_j|$. Based on the result above, the convergence time $t(\varepsilon)$ after which total variation falls below ε (i.e. $\forall k \geq t(\varepsilon) : \Delta(\boldsymbol{\pi}_k, \boldsymbol{\pi}) \leq \varepsilon$) is than given as

$$t(\varepsilon) = \frac{1}{\ln(|\lambda_2|)} \cdot \left(\ln(\varepsilon) - \ln\left(\frac{1}{2} \sum_{j=1}^{n} |a_2 \left(\mathbf{v}_2\right)_j|\right)\right).$$

We now consider the null model $\tilde{\mathbf{T}}^{(2)}$ corresponding to a Markovian temporal network model derived from $G^{(2)}$ (and thus to a random walk running on the weighted aggregate network) according to Eq. 3 in our main article. Based on the sequence of eigenvalues $1 = \tilde{\lambda}_1 \geq |\tilde{\lambda}_2| \geq \ldots \geq |\tilde{\lambda}_n|$ of $\tilde{\mathbf{T}}^{(2)}$ with corresponding eigenvectors $\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n$, a convergence time $\tilde{t}(\varepsilon)$ after which total variation distance falls below ε can then be derived analogously as:

$$\tilde{t}(\varepsilon) = \frac{1}{\ln(|\tilde{\lambda}_2|)} \cdot \left(\ln(\varepsilon) - \ln\left(\frac{1}{2} \sum_{j=1}^{n} |\tilde{a}_2 \left(\tilde{\mathbf{v}}_2\right)_j|\right)\right)$$

A fraction $\mathcal{S}^*(\mathbf{T}^{(2)}, \varepsilon)$ that captures the slow-down (or speed-up) of convergence that is

due to non-Markovian properties can then be defined based on $t(\varepsilon)/\tilde{t}(\varepsilon)$:

$$\mathcal{S}^*(\mathbf{T}^{(2)},\varepsilon) := \frac{\ln(|\tilde{\lambda}_2|)}{\ln(|\lambda_2|)} \cdot \frac{\ln(\varepsilon) - \ln\left(\frac{1}{2}\sum_{j=1}^{n} |a_2\,(\mathbf{v}_2)_j\,|\right)}{\ln(\varepsilon) - \ln\left(\frac{1}{2}\sum_{j=1}^{n} |\tilde{a}_2\,(\tilde{\mathbf{v}}_2)_j\,|\right)}$$

We then define the proportional slow-down $\mathcal{S}^*(\mathbf{T}^{(2)})$ in the limit of small ε (or large k) as

$$\mathcal{S}^*(\mathbf{T}^{(2)}) := \lim_{\varepsilon\to 0}\left(\mathcal{S}^*(\mathbf{T}^{(2)},\varepsilon)\right) = \frac{\ln(|\tilde{\lambda}_2|)}{\ln(|\lambda_2|)}. \tag{A.4}$$

We remark, that this slow-down is due to the difference in the spectral gap $1-|\lambda_2|$ of $\mathbf{T}^{(2)}$ as compared to the null-model $\tilde{\mathbf{T}}^{(2)}$ derived from the weighted aggregate network corresponding to both $\mathbf{T}^{(2)}$ and $\tilde{\mathbf{T}}^{(2)}$. The prediction $\mathcal{S}^*(\mathbf{T}^{(2)})$ holds for sufficiently large k or - equivalently - for a sufficiently small total variation distance ε. Furthermore, we assumed that $\tilde{\mathbf{T}}^{(2)}$ is primitive and that λ_2 is non-degenerate.

If the gap $1-|\tilde{\lambda}_2|$ of the second-order network corresponding to the Markovian temporal network is larger than the gap $1-|\lambda_2|$ corresponding to a non-Markovian case, $\mathcal{S}^*(\mathbf{T}^{(2)}) > 1$. In this case, the *conductance* of $\tilde{G}^{(2)}$ is larger than that of $G^{(2)}$ and the non-Markovian properties slow down random walk convergence. If - on the other hand - the gap $1-|\tilde{\lambda}_2|$ is smaller than the gap $1-|\lambda_2|$, the conductance of $\tilde{G}^{(2)}$ is smaller than that of $G^{(2)}$. In this case $\mathcal{S}^*(\mathbf{T}^{(2)}) < 1$, meaning that the non-Markovian properties of a temporal network speed up random walk convergence.

We finally note that for $|\lambda_2| = |\lambda_3|$, a similar slow-down ratio can be derived for the chi-square distance based on the upper bounds on the second-largest eigenvalues for general directed networks with arbitrary eigenvalue spectra following the arguments put forth in [27]. Based on this approach the prediction would look like

$$S^*_\chi(\mathbf{T}^{(2)}) = \frac{\ln\left(\frac{1}{2}(1+\mathrm{Re}(\tilde{\lambda}_2))\right)}{\ln\left(\frac{1}{2}(1+\mathrm{Re}(\lambda_2))\right)},$$

with the eigenvalue sequence of the transition matrix sorted by their real parts, i.e. $\mathrm{Re}(\lambda_1) \geq \mathrm{Re}(\lambda_2) \geq \ldots \geq \mathrm{Re}(\lambda_n)$. The prediction $S^*_\chi(\mathbf{T}^{(2)})$ isa equal to $S^*(\frac{1}{2}(\mathbf{I}_n + \mathbf{T}^{(2)}))$ where n is the dimension of $\mathbf{T}^{(2)}$ and $\mathbf{I}_n$ is the corresponding identity matrix. This is equal to applying the prediction S^* to a transition matrix of a lazy random walk with self-loop probability 1/2. This approach can alleviate periodicity and assure that $|\lambda_2| > |\lambda_3|$ at least for the transition matrix of a lazy random walk.

A.2 Details of model for non-markovian temporal networks

A particularly important finding in our article is the fact that non-Markovian characteristics can give rise both to a slow-down and speed-up of diffusion dynamics when compared to their static aggregated counterparts. To illustrate this fact, we introduce a simple toy model for temporal networks in which non-Markovian properties can either *inhibit* or *enforce* time-respecting paths across two pronounced communities that are present in the static aggregate network. In our article we argue that the presence of order correlations which enforce time-respecting paths across communities is a particularly simple mechanism by which non-Markovian properties in temporal networks can speed up diffusion dynamics. With this we further highlight one possible mechanism by which non-Markovian properties can effectively mitigate the decelerating effect of community structures on diffusion dynamics.

In the following, we formally define our toy model and substantiate our interpretations in the article by means of a spectral analysis of the second-order aggregate networks corresponding to different points in the model's parameter space. The model is based on a directed, weighted aggregate network $G^{(1)}$ with two communities, each consisting of a random k-regular graph with n nodes. To interconnect the two communities, we randomly draw links $e = (v_1, v_2)$ and $e' = (v_1', v_2')$ from the two communities respectively, remove e and e' and instead add links (v_1, v_1') and (v_2, v_2') thus maintaining a k-regular aggregate network. We further assign uniform weights ω_1 to all links, thus obtaining a network as shown in the schematic illustration in panel (a) of Supplementary Figure A.1. For the simulations in the article, we use $k = 4$ and $n = 50$, thus obtaining a network with 100 nodes and 400 directed links.

For this first-order network $G^{(1)}$, we construct a second-order network $G^{(2)}$ corresponding to Markovian link activations as shown in panel (b) of Supplementary Figure A.1. Since $G^{(1)}$ has 400 links, $G^{(2)}$ has 400 nodes, each corresponding to a directed link in the first-order network. As weights in the second-order network $G^{(2)}$, we consider a uniform constant ω_2 which corresponds to a Markovian case in which consecutive link activations are independently drawn. We use the following simple strategy to introduce non-Markovian properties. We first identify all links (x, y) that interconnect the two communities, i.e. where x is a node in community 1 and y is a node in community 2. For these links, we then identify two nodes a, b such that a is a node in community 1 adjacent to x and b is a node in community 2 adjacent to node y. The basic idea of the model is to change the weights of those two-paths that involve links $(a, x), (x, y), (y, x)$ and (x, a). The statistics

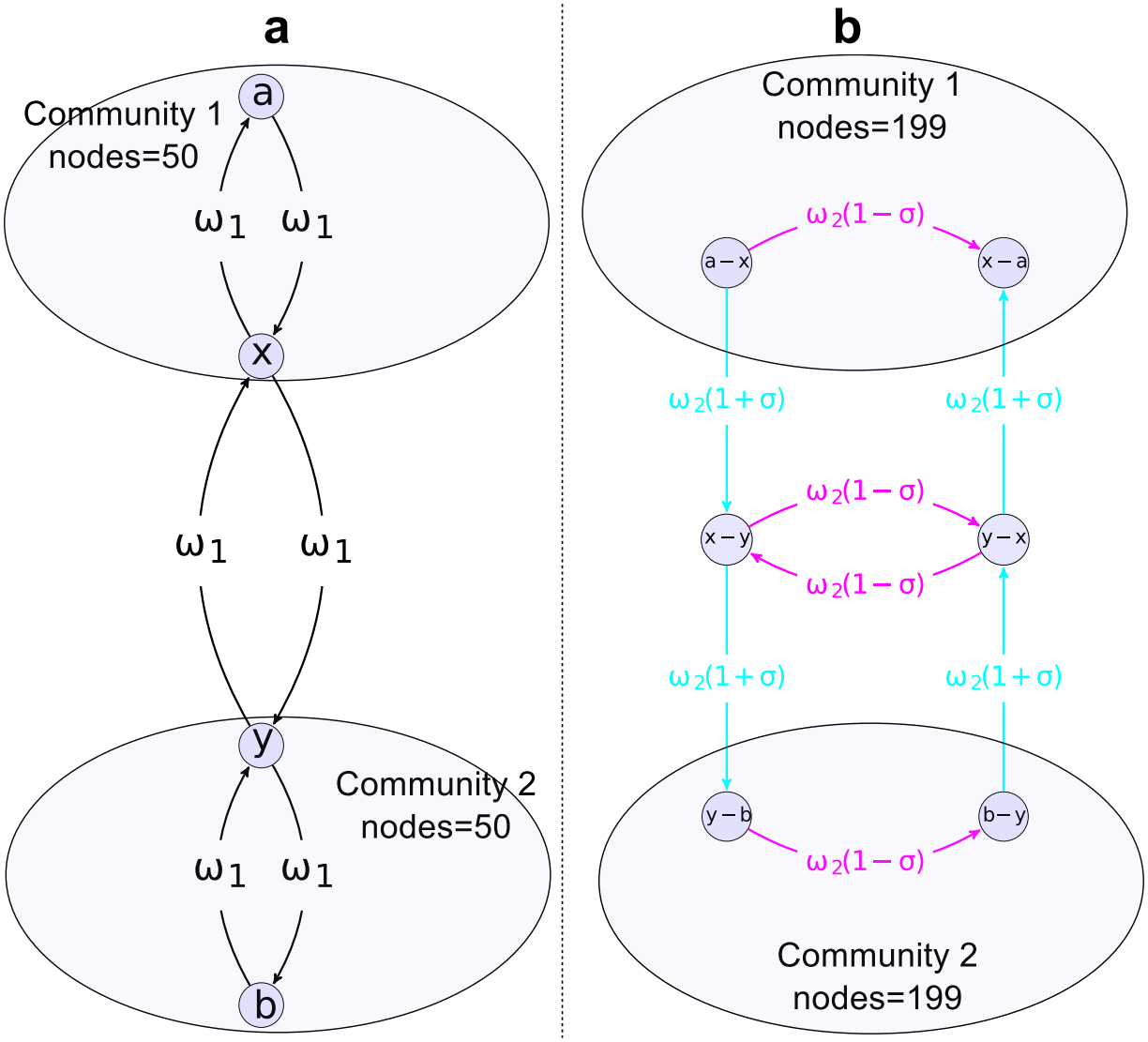


Figure A.1: Schematic representation of our model for non-Markovian temporal networks (a) The first-order aggregate network $G^{(1)}$ consists of two pronounced communities connected by directed inter-community links (x, y) and (y, x). (b) Weights in the corresponding second-order aggregate network $G^{(2)}$ are changed by means of a parameter σ. Positive values for σ enforce two-paths across communities (turquoise) and inhibit two-paths within communities (magenta).

of these two-paths is captured by the weights of *links* connecting *nodes* (a, x), (x, y), (x, a), (a, x) in the second-order network (see panel (b) in Supplementary Figure A.1).

Based on a parameter $\sigma \in (-1, 1)$, the weights of the second-order links $(a, x) \to (x, y)$ and $(y, x) \to (x, a)$ are set to $\omega_2(1 + \sigma)$, while the weights of second-order links $(a, x) \to (x, a)$ and $(y, x) \to (x, y)$ are set to $\omega_2(1 - \sigma)$. Weights of second-order links including the nodes

b and y are adjusted analogously (see panel (b) in Supplementary Figure A.1). By this means, positive values for σ increase the weights of two-paths *across communities* at the expense of two-paths *within communities*. Negative values for σ increase the weights of two-paths *within communities* at the expense of two-paths *across communities*. A value of $\sigma = 0$ yields a second-order aggregate network with uniform weights ω_2 which - by construction - corresponds to a Markovian case.

For $\sigma \neq 0$, the above procedure leads to transition matrices $\mathbf{T}^{(2)} \neq \tilde{\mathbf{T}}^{(2)}$ which are however consistent with the same weighted aggregate network $G^{(1)}$. This can be confirmed by checking that for all $\sigma \in (-1, 1)$, the stationary activation frequencies of links captured by the leading eigenvector $\boldsymbol{\pi}$ of $\mathbf{T}^{(2)}$ are the same. The change of second-order weights by our model imply

$$\begin{aligned} T^{(2)}_{(a,x)(x,y)} &= \omega_2(1+\sigma), \quad T^{(2)}_{(y,x)(x,a)} = \omega_2(1+\sigma)\,, \\ T^{(2)}_{(a,x)(x,a)} &= \omega_2(1-\sigma), \quad T^{(2)}_{(y,x)(x,y)} = \omega_2(1-\sigma)\,. \end{aligned}$$

Since the j-th component of the stationary distribution of the second-order network is given by $(\boldsymbol{\pi})_j = \sum_i (\boldsymbol{\pi})_i T^{(2)}_{ij}$ the changes above only influence entries $(\boldsymbol{\pi})_{(x,a)}$ and $(\boldsymbol{\pi})_{(x,y)}$ in the leading eigenvector of $\mathbf{T}^{(2)}$. Let $\tilde{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}}\tilde{\mathbf{T}}^{(2)}$ and $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{T}^{(2)}$. Then for an entry $(\boldsymbol{\pi})_{(x,a)}$ we can write

$$\begin{aligned} (\boldsymbol{\pi})_{(x,a)} &= \sum_i (\boldsymbol{\pi})_{(i,x)} T^{(2)}_{(i,x)(x,a)} \\ &= \sum_{i \neq a,y} \left((\boldsymbol{\pi})_{(i,x)} T^{(2)}_{(i,x)(x,a)} \right) \\ &\quad + (\boldsymbol{\pi})_{(a,x)} T^{(2)}_{(a,x)(x,a)} + (\boldsymbol{\pi})_{(y,x)} T^{(2)}_{(y,x)(x,a)}\,. \end{aligned}$$

Recall that our transformations only change the entries for (x,a) and (x,y) therefore it holds that $(\boldsymbol{\pi})_{(i,x)} = (\tilde{\boldsymbol{\pi}})_{(i,x)}$ for all i. This yields

$$\begin{aligned} (\boldsymbol{\pi})_{(x,a)} &= \sum_{i \neq a,y} \left((\tilde{\boldsymbol{\pi}})_{(i,x)} T^{(2)}_{(i,x)(x,a)} \right) \\ &\quad + (\tilde{\boldsymbol{\pi}})_{(a,x)} T^{(2)}_{(a,x)(x,a)} + (\tilde{\boldsymbol{\pi}})_{(y,x)} T^{(2)}_{(y,x)(x,a)}\,. \end{aligned}$$

Furthermore, we can plug in the definitions for $\mathbf{T}^{(2)}$ from above and also use that $T^{(2)}_{(i,x)(x,a)} =$

$\tilde{T}^{(2)}_{(i,x)(x,a)}$ for all $i \notin \{a, y\}$.

$$\begin{aligned}
(\boldsymbol{\pi})_{(x,a)} &= \sum_{i \neq a,y} \left((\tilde{\boldsymbol{\pi}})_{(i,x)} \, \tilde{T}^{(2)}_{(i,x)(x,a)} \right) + (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \omega_2 (1 - \sigma) \\
&\quad + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \omega_2 (1 + \sigma) \\
&= \sum_{i \neq a,y} \left((\tilde{\boldsymbol{\pi}})_{(i,x)} \, \tilde{T}^{(2)}_{(i,x)(x,a)} \right) \\
&\quad + (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \omega_2 - (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \omega_2 \sigma \\
&\quad + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \omega_2 + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \omega_2 \sigma \, .
\end{aligned}$$

Since $\tilde{\mathbf{T}}^{(2)}$ is built from a regular graph it holds that $\omega_2 = \tilde{T}^{(2)}_{(i,x)(x,a)}$ for all i. Hence,

$$\begin{aligned}
(\boldsymbol{\pi})_{(x,a)} &= \sum_{i \neq a,y} \left((\tilde{\boldsymbol{\pi}})_{(i,x)} \, \tilde{T}^{(2)}_{(i,x)(x,a)} \right) + (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \tilde{T}^{(2)}_{(a,x)(x,a)} \\
&\quad - (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \omega_2 \sigma + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \tilde{T}^{(2)}_{(y,x)(x,a)} + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \omega_2 \sigma \\
&= \sum_{i} \left((\tilde{\boldsymbol{\pi}})_{(i,x)} \, \tilde{T}^{(2)}_{(i,x)(x,a)} \right) \\
&\quad - (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \omega_2 \sigma + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \omega_2 \sigma \\
&= (\tilde{\boldsymbol{\pi}})_{(x,a)} - (\tilde{\boldsymbol{\pi}})_{(a,x)} \, \omega_2 \sigma + (\tilde{\boldsymbol{\pi}})_{(y,x)} \, \omega_2 \sigma \\
&= (\tilde{\boldsymbol{\pi}})_{(x,a)} \, .
\end{aligned}$$

In the last step we use that the stationary distribution $\tilde{\boldsymbol{\pi}}$ is uniform and thus $(\tilde{\boldsymbol{\pi}})_{(a,x)} = (\tilde{\boldsymbol{\pi}})_{(y,x)}$. From an analogous argumentation, we can derive $(\boldsymbol{\pi})_{(x,y)} = (\tilde{\boldsymbol{\pi}})_{(x,y)}$. We thus confirm that $\boldsymbol{\pi} = \tilde{\boldsymbol{\pi}}$ and the stationary distribution is preserved for $\sigma \in (-1, 1)$. We finally refer the reader to a related model for non-Markovian temporal networks, which has been introduced very recently, during the revision of our manuscript [86]. Different from our approach, the model introduced in this recent work generates realisations that do not preserve a given weighted aggregate network, which however is the particular focus of our approach.

Appendix B

Interconnectivity

B.1 Proofs and derivations

Note: Unless stated otherwise, here vectors are considered to be row-vectors and multiplication of vectors with matrices are left multiplications.

We assume a multi-layer network $\mathbf{G}$ consisting of L layers $G_1, \ldots, G_L$ and n nodes. A single layer G_s contains n_s nodes and therefore $\sum_{s=1}^{L} n_s = n$. For a multi-layer network $\mathbf{G}$ we define the *supra-transition* matrix that can be represented in block structure according to the layers:

$$\mathbf{T} = \left(\begin{array}{c|c|c|c|c} \mathbf{T}_1 & \ldots & \mathbf{T}_{1t} & \ldots & \mathbf{T}_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \mathbf{T}_{s1} & \ldots & \mathbf{T}_{st} & \ldots & \mathbf{T}_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \mathbf{T}_{L1} & \ldots & \mathbf{T}_{Lt} & \ldots & \mathbf{T}_{L} \end{array}\right) .$$

Each $\mathbf{T}_{st}$ contains all the transition probabilities from nodes in G_s to nodes in G_t. Assuming Eq.(6.5) it follows that $\mathbf{T}_{st} = \alpha_{st}\mathbf{R}_{st}$ where $\mathbf{R}_{st}$ is a row stochastic matrix. This means that all $\mathbf{T}_{st}$ are scaled transition matrices. The factor α_{st} represents the weighted fraction of all links starting in G_s that end up in G_t.

In this respect we define the aggregated transition matrix $\mathfrak{T}$ of dimension L,

$$\mathfrak{T} = \left(\begin{array}{c|c|c|c|c} \alpha_{11} & \dots & \alpha_{1t} & \dots & \alpha_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \alpha_{s1} & \dots & \alpha_{st} & \dots & \alpha_{sL} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline \alpha_{L1} & \dots & \alpha_{Lt} & \dots & \alpha_{L} \end{array}\right) . \tag{B.1}$$

Each vector v of dimension n can be split according to the layer-separation given by $\mathbf{G}$,

$$v = \left(v^{(1)}, \dots, v^{(k)}, \dots, v^{(L)}\right) .$$

Each component $v^{(k)}$ has exactly dimension n_k. We define the *layer-aggregated vector* $\mathfrak{v} = (\mathfrak{v}_1, \dots, \mathfrak{v}_L)$ of dimension L as follows

$$\forall k \in \{1, \dots, L\} \quad \mathfrak{v}_k = \sum_{i=1}^{n_k} \left[v^{(k)}\right]_i .$$

We use the bracket notation $[v]_i$ to represent the i-th entry of the vector v. Analogously, by $[v\mathbf{M}]_i$ we mean the i-th entry w_i that represents the multiplication of v with a matrix $\mathbf{M}$, i.e. $w = v\mathbf{M}$. Further, by $|v|$ we indicate the sum of the entries of v, $|v| = \sum_i v_i = \sum_i [v]_i$.

Theorem 1. *For a multi-layer network $\mathbf{G}$ consisting of L layers we assume the supra-transition matrix $\mathbf{T}$ to consist of block matrices $\mathbf{T}_{st}$ such that for all $s, t \in \{1, \dots, L\}$, $\mathbf{T}_{st} = \alpha_{st}\mathbf{R}_{st}$ where $\alpha_{st} \in \mathbb{Q}$ and $\mathbf{R}_{st}$ is a row stochastic transition matrix. The multi-layer aggregation is defined by $\mathfrak{T} = \{\alpha_{st}\}_{st}$. If an eigenvalue λ of the matrix $\mathbf{T}$ corresponds to an eigenvector v with a layer-aggregation $\mathfrak{v}$ that satisfies $\mathfrak{v} \neq 0$ then λ is also an eigenvalue of $\mathfrak{T}$.*

Proof. Assume v is a left eigenvector of $\mathbf{T}$ corresponding to the eigenvalue λ. Therefore, it holds that $v\mathbf{T} = \lambda v$. Let $v^{(k)}$ be the k-th part of v that corresponds to the layer G_k. We can write the matrix multiplication in block structure.

$$\left(v^{(1)}, \dots, v^{(k)}, \dots, v^{(L)}\right)\mathbf{T} = \left(\sum_l v^{(l)}\mathbf{T}_{l1}, \dots, \sum_l v^{(l)}\mathbf{T}_{lk}, \dots, \sum_l v^{(l)}\mathbf{T}_{lL}\right) .$$

Each $v^{(k)}$ is a row vector which length is equal to the amount of nodes n_k in G_k. The transformation $\sum_l v^{(l)}\mathbf{T}_{lk}$ is also a row vector with the same length as $v^{(k)}$. According to

the eigenvalue equation it holds that for all $k \in \{1, \dots, L\}$

$$\lambda v^{(k)} = (v\mathbf{T})^{(k)} = \sum_l v^{(l)} \mathbf{T}_{lk} \,.$$

Now let us denote the sum of the vector entries of $v^{(k)}$ as

$$\mathfrak{v}_k = \sum_i \left[v^{(k)}\right]_i \,.$$

Further, we define layer-aggregated vector consisting of this sums by $\mathfrak{v} = (\mathfrak{v}_1, \dots, \mathfrak{v}_n)$. Note that for a general row stochastic matrix $\mathbf{M}$ and its multiplication with any vector v tit holds that $\sum_j [v]_j = \sum_j [v\mathbf{M}]_j$. For the components after multiplication with $\mathbf{T}$ we can deduce

$$\begin{aligned}\sum_i \left[(v\mathbf{T})^{(k)}\right]_i &= \sum_i \left[\sum_l v^{(l)} \mathbf{T}_{lk}\right]_i = \sum_i \left[\sum_l \alpha_{lk} v^{(l)} \mathbf{R}_{lk}\right]_i \\ &= \sum_l \alpha_{lk} \sum_i \left[v^{(l)} \mathbf{R}_{lk}\right]_i = \sum_l \alpha_{lk} \sum_i \left[v^{(l)}\right]_i = \sum_l \alpha_{lk} \mathfrak{v}_l \,.\end{aligned}$$

If we multiply $\mathfrak{v}$ with $\mathfrak{T}$ and look at a single entry of $\mathfrak{v}\mathfrak{T}$ we get

$$[\mathfrak{v}\mathfrak{T}]_k = \sum_l \mathfrak{v}_l \mathfrak{T}_{lk} = \sum_l \mathfrak{v}_l \alpha_{lk} \,.$$

Hence it holds that

$$[\mathfrak{v}\mathfrak{T}]_k = \sum_i \left[(v\mathbf{T})^{(k)}\right]_i \,,$$

and therefore

$$\mathfrak{v}\mathfrak{T} = \left(\sum_i \left[(v\mathbf{T})^{(1)}\right]_i, \dots, \sum_i \left[(v\mathbf{T})^{(L)}\right]_i\right) \,.$$

Finally since $\mathbf{T}$ is row stochastic and $\lambda v = v\mathbf{T}$ we have that

$$\begin{aligned}\lambda\mathfrak{v} &= \lambda \left(\mathfrak{v}_1, \dots, \mathfrak{v}_n\right) \\ &= \lambda \left(\sum_i [v^{(1)}]_i, \dots, \sum_i [v^{(L)}]_i\right) \\ &= \left(\sum_i \left[(v\mathbf{T})^{(1)}\right]_i, \dots, \sum_i \left[(v\mathbf{T})^{(L)}\right]_i\right) \\ &= \mathfrak{v}\mathfrak{T} \,.\end{aligned}$$

Therefore, λ is also an eigenvalue of $\mathfrak{T}$ to the eigenvector $\mathfrak{v}$ defined as before. It is important to note that this only holds if $\mathfrak{v} \neq 0$. □

The procedure used in the proof of the previous theorem applies to several eigenvalues of $\mathbf{T}$ but at most L of them. Next we give a proposition for the reversed statement of Thm 1.

Proposition 1. *Let* $\mathbf{G}$ *be a multi-layer network that consists of* L *layers and fulfills all of the conditions of Thm 1. Let* $\mathfrak{T} = \{\alpha_{st}\}_{st}$ *be the multi-layer aggregation of* $\mathbf{T}$*. If* λ *is an eigenvalue of* $\mathfrak{T}$ *then* λ *is also an eigenvalue of* $\mathbf{T}$*.*

Proof. Assume that λ is an eigenvalue of $\mathfrak{T}$. For each matrix there exist a left and right eigenvector that correspond to the same eigenvalue λ. Assume the $\mathfrak{w}$ is the right eigenvector and therefore a column vector. Hence $\mathfrak{T}\mathfrak{w} = \lambda\mathfrak{w}$ and

$$\mathfrak{T}\mathfrak{w} = \left(\sum_j \alpha_{1j}\mathfrak{w}_j, \ldots, \sum_j \alpha_{kj}\mathfrak{w}_j, \ldots, \sum_j \alpha_{Lj}\mathfrak{w}_j\right)^\top = \lambda\mathfrak{w}\,. \tag{B.2}$$

Now we generate a column vector w of dimension n such that for all the layer components $w^{(k)}$ it holds that

$$w^{(k)} = (\mathfrak{w}_k, \ldots, \mathfrak{w}_k)^\top\,. \tag{B.3}$$

Next we perform a right multiplication of w with $\mathbf{T}$,

$$\begin{aligned}\mathbf{T}w &= \left(\sum_l \mathbf{T}_{1l}w^{(l)}, \ldots, \sum_l \mathbf{T}_{kl}w^{(l)}, \ldots, \sum_l \mathbf{T}_{Ll}w^{(l)}\right)^\top \\ &= \left(\sum_l \alpha_{1l}\mathbf{R}_{1l}w^{(l)}, \ldots, \sum_l \alpha_{kl}\mathbf{R}_{kl}w^{(l)}, \ldots, \sum_l \alpha_{Ll}\mathbf{R}_{Ll}w^{(l)}\right)^\top.\end{aligned}$$

Since all $\mathbf{R}_{st}$ are row stochastic matrices and $w^{(l)}$ contains only the value $\mathfrak{w}_l$ for each entry we get $\mathbf{R}_{st}w^{(l)} = w^{(l)}$. It follows that

$$\begin{aligned}\mathbf{T}w &= \left(\sum_l \alpha_{1l}w^{(l)}, \ldots, \sum_l \alpha_{kl}w^{(l)}, \ldots, \sum_l \alpha_{Ll}w^{(l)}\right)^\top \\ &= \left(\lambda w^{(1)}, \ldots, \lambda w^{(k)}, \ldots, \lambda w^{(L)}\right)^\top \\ &= \lambda w\,.\end{aligned}$$

And therefore λ is also an eigenvalue of $\mathbf{T}$. □

In the case of a diffusion process we are especially interested in the second-largest eigenvalue of $\mathbf{T}$, denoted by $\lambda_2(\mathbf{T})$, which is related to algebraic connectivity of $\mathbf{T}$. In this perspective the following corollary is useful:

Corollary 1. *Let* $\mathbf{G}$ *be a multi-layer network consisting of* L *layers that fulfill all of the conditions of Thm 1. Further assume that* $\mathbf{G}$ *is partitioned according to a spectral partitioning, i.e. according to the eigenvector corresponding to* $\lambda_2(\mathbf{T})$*, then* $\lambda_2(\mathbf{T}) = \lambda_2(\mathfrak{T})$*.*

Proof. In general all the eigenvectors of a transition matrix, except the eigenvector corresponding to the largest eigenvalue that is equal to one, sum up to zero. However, these eigenvectors consist of positive and negative entries that allow for a spectral partitioning. Especially the eigenvector v_2 that corresponds to the second-largest eigenvalue $\lambda_2(\mathbf{T})$, can be used for the partitioning of the network. This eigenvector is related to the Fiedler vector that is also used for spectral bisection [38]. Therefore if the layer-partition of $\mathbf{G}$ coincides with this spectral partitioning we assure that the layer-aggregated vector of v_2 satisfies $\mathfrak{v}_2 \neq 0$. Considering this and Prop 1 the corollary follows directly from Thm 1. □

Given Eq.(6.5) we can fully describe the spectrum of $\mathbf{T}$ based on the intra-layers transition blocks $\mathbf{T}_i$ for $i \in \{1, \ldots, n\}$ and the spectrum of $\mathfrak{T}$. Note that with uniform columns of a matrix $\mathbf{M}$ we mean that each column of $\mathbf{M}$ contains the same value in each entry. However, this value can be different for different columns.

Proposition 2. *Let* $\mathbf{T}$ *be the supra-transition matrix of a multi-layer network* $\mathbf{G}$ *that consist of* L *layers and satisfies Eq.(6.5). If* $\mathbf{T}$ *has off-diagonal block matrices* $\mathbf{T}_{st}$*, for* $s, t \in \{1, \ldots, n\}$ *and* $s \neq t$*, that all have uniform columns, then the spectrum of* $\mathbf{T}$ *can be decomposed as*

$$Spec(\mathbf{T}) = \{1, \lambda_2, \ldots, \lambda_L\} \cup \left(\bigcup_{s=1}^{L} Spec(\mathbf{T}_s) \setminus \{\lambda_1(\mathbf{T}_s)\} \right), \tag{B.4}$$

where $\mathbf{T}_s$ *are the block matrices of* $\mathbf{T}$ *corresponding to the single layers* G_s *and* $\lambda_1(\mathbf{T}_s)$ *the largest eigenvalue of* $\mathbf{T}_s$*. The eigenvalues* $\lambda_2, \ldots, \lambda_L$ *are attributed to the interconnectivity of layers.*

Proof. To prove this statement we just have to show that all eigenvalues (except the largest one) of $\mathbf{T}_s$ for $s \in \{1, \ldots, L\}$ are also eigenvalues of $\mathbf{T}$. We assume that λ is any eigenvalue corresponding to the eigenvector u of some block matrix $\mathbf{T}_r$, i.e. $\lambda u = u\mathbf{T}_r$. We define a row vector v that is zero everywhere except at the position where it corresponds to $\mathbf{T}_r$.

The vector v looks like $v = (0, \ldots, 0, u, 0, \ldots, 0)$. Now we investigate what happens if we multiply this vector with the transition matrix $\mathbf{T}$.

$$v\mathbf{T} = \left(v^{(1)}, \ldots, v^{(k)}, \ldots, v^{(L)}\right)\mathbf{T} = \left(\sum_l v^{(l)}\mathbf{T}_{l1}, \ldots, \sum_l v^{(l)}\mathbf{T}_{lk}, \ldots, \sum_l v^{(l)}\mathbf{T}_{lL}\right) .$$

Let us take a look at the effect of the matrix multiplication on an arbitrary component $v^{(k)}$ with $k \neq r$ and recall that $v^{(k)}$ is equal to a zero vector $\mathbf{0}$ for $k \neq r$.

$$(v\mathbf{T})^{(k)} = \sum_l v^{(l)}\mathbf{T}_{lk} = \sum_{l,l\neq r} \mathbf{0}\mathbf{T}_{lk} + u\mathbf{T}_{rk} = u\mathbf{T}_{rk} .$$

Note that all eigenvectors u of a transition matrix that are not related to the largest eigenvalue sum up to zero. Therefore it holds that $u\mathbf{T}_{rk} = \mathbf{0}$ since $\mathbf{T}_{rk}$ has uniform columns and therefore $u\mathbf{T}_{rk}$ yields in a vector where each entry is equal to some multiple of $|u|$. In case of $k = r$ it holds that $v^{(k)} = u$ and we get

$$(v\mathbf{T})^{(r)} = \sum_l v^{(l)}\mathbf{T}_{lr} = \sum_{l,l\neq r} \mathbf{0}\mathbf{T}_{lk} + u\mathbf{T}_{rr} = u\mathbf{T}_r = \lambda r .$$

Hence, it holds that $v\mathbf{T} = \lambda v$, which means that λ is also an eigenvalue of $\mathbf{T}$. This way we get $n - L$ eigenvalues of $\mathbf{T}$ apart from the largest eigenvalue that is equal to one. The remaining eigenvalues denoted by $\lambda_2, \ldots, \lambda_L$ are not attributed to any block matrix of $\mathbf{T}$. Therefore they are considered to be the interconnectivity eigenvalues. □

Corollary 2. *Let* $\mathbf{G}$ *be a multi-layer network consisting of L layers that satisfies Eq.(6.5) and the conditions of Prop 2. Then the aggregated matrix* $\mathfrak{T} = \{\alpha_{st}\}_{st}$ *has spectrum*

$$Spec(\mathfrak{T}) = \{1, \lambda_2, \ldots, \lambda_L\} ,$$

and it holds that $\lambda_2, \ldots, \lambda_L \in Spec(\mathbf{T})$.

Proof. Note that every eigenvalue $\lambda \neq 1$ of some block matrix $\mathbf{T}_r$ with $\lambda u = u\mathbf{T}_r$ is by Prop 2 also an eigenvalue of $\mathbf{T}$. Furthermore, λ is attributed to the eigenvector $v = (0, \ldots, 0, u, 0, \ldots, 0)$ of $\mathbf{T}$. However $|v| = 0$ since u is an eigenvector of a transition matrix, not related to the largest eigenvalue, and therefore sums up to zero. Hence all eigenvalues fulfilling this condition are by Thm 1 not eigenvalues of $\mathfrak{T}$. Since $\mathfrak{T}$ contains at least L eigenvalues that by Prop 1 also correspond to eigenvalues of $\mathbf{T}$, the remaining eigenvalues $\lambda_2, \ldots, \lambda_L$ have to also be eigenvalues of $\mathfrak{T}$. □

In the following we provide a useful proposition for the eigenvalues arising from the inter-

links in case of two layers. Note that by the function $T(\cdot)$ applied to matrix $\mathbf{M}$ we indicate that $T(\mathbf{M})$ is the row-normalization of $\mathbf{M}$.

Proposition 3. *Let* $\mathbf{G}$ *be a multi-layer network that satisfies Eq.(6.5), consisting of two networks* G_1 *and* G_2 *in separate layers. Assume that the supra-transition matrix* $\mathbf{T}$ *has the form*

$$\mathbf{T} = \left(\begin{array}{c|c} \mathbf{T}_1 & \mathbf{T}_{12} \\ \hline \mathbf{T}_{21} & \mathbf{T}_2 \end{array}\right) = \left(\begin{array}{c|c} \beta T(\mathbf{A}_1) & (1-\beta)\mathbf{T}_{12}^I \\ \hline (1-\beta)\mathbf{T}_{21}^I & \beta T(\mathbf{A}_2) \end{array}\right),$$

where $\mathbf{T}^I$ *is the transition matrix of the layer* $\mathbf{G}$ *that only consists of the inter-layer links and* $\beta \in \mathbb{Q}$ *is a constant. Furthermore, assume that* $\mathbf{T}_1$ *and* $\mathbf{T}_2$ *have uniform columns. Then for* $\lambda \in Spec(\mathbf{T}^I)$ *and* $\lambda \neq 1, -1$ *it holds that* $(1-\beta)\lambda \in Spec(\mathbf{T})$.

Proof. If v is an eigenvector to the eigenvalue $\lambda \neq 1, -1$ of $\mathbf{T}^I$ it holds that $v\mathbf{T}^I = \lambda v$. Hence,

$$v\mathbf{T}^I = \left(v^{(1)}, v^{(2)}\right)\mathbf{T}^I = \left(v^{(2)}\mathbf{T}_{21}^I, v^{(1)}\mathbf{T}_{12}^I,\right) = \lambda\left(v^{(1)}, v^{(2)}\right).$$

Because $\lambda v^{(2)} = v^{(1)}\mathbf{T}_{12}^I$, we get $\lambda^2 v^{(1)} = v^{(1)}\mathbf{T}_{12}^I\mathbf{T}_{21}^I$. Therefore, $v^{(1)}$ is also an eigenvector of the transition matrix $\mathbf{T}_{12}^I\mathbf{T}_{21}^I$ to the eigenvalue λ^2. Note that $\lambda \neq 1, -1$ hence $\lambda^2 < 1$ which implies that $v^{(1)}$ does not correspond to the largest eigenvalue and therefore its entries sum up to zero. The same holds for $v^{(2)}$ and the matrix $\mathbf{T}_{21}^I\mathbf{T}_{12}^I$. For the multiplication of v with $\mathbf{T}$ we deduce that

$$v\mathbf{T} = \left(v^{(1)}, v^{(2)}\right)\mathbf{T} = \left(v^{(1)}\mathbf{T}_1 + (1-\beta)v^{(2)}\mathbf{T}_{21}^I, (1-\beta)v^{(1)}\mathbf{T}_{12}^I + v^{(2)}\mathbf{T}_2,\right).$$

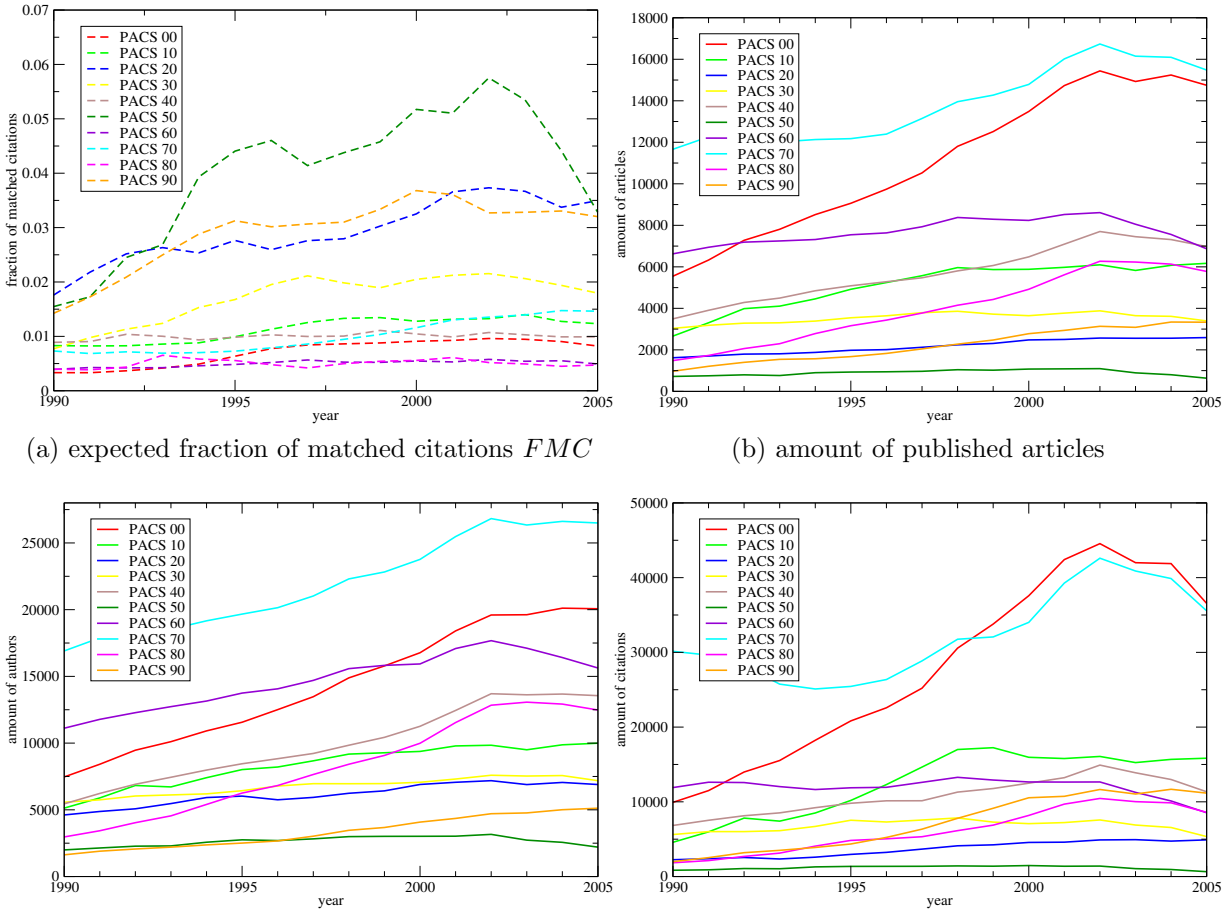
Since $\mathbf{T}_1$ and $\mathbf{T}_2$ have uniform columns we get $v^{(1)}\mathbf{T}_1 = \mathbf{0}$ and $v^{(2)}\mathbf{T}_2 = \mathbf{0}$. And therefore $v\mathbf{T} = (1-\beta)\lambda v$ and $(1-\beta)\lambda \in Spec(\mathbf{T})$. □

Proposition 3 can be extended to multiple layers, however the proof is more involved and will be omitted.

B.2 Scientometric data

Here we provide additional plots corresponding to the multi-layer analysis performed in Chapter 7. The observations are based on the APS data set [1] during time periods of 5 year with starting years ranging from 1990 to 2005. In Figure B.1 we depict the time evolution several properties of the article citation network. The expected amount of matched citations in Figure B.1 (a) was already discussed in the corresponding chapter. However, in the presented scale we can see that expected FMC values follow similar trends

than the real FMC values. Therefore, the difference between PACS classes are most likely due to network statistics.



(a) expected fraction of matched citations FMC

(b) amount of published articles

(c) amount of authors that published an article

(d) amount of total citations

Figure B.1: Time evolution of several network properties The y-axis shows the values of several network proeprties for sliding time-windows of 5 years. The x-axis indicate the starting year of a time-window of 5 years. Only articles published during this period are considered for the analysis. The main PACS classes are represented by different colors.

Further, we can see that the relative time-evolution of published articles and amount of authors involved in these articles are quite similar an increase over time. The amount of total citations depict in Figure B.1 (d) also correlate with the size of the system according to articles and authors. PACS classes 00, 20 and 50 are the smallest systems but exhibit the largest expected value of matched citations. Hence, smaller system size seems to foster more co-authorships in comparison to the amount of citations.

Bibliography

[1] (2013). APS Data Sets for Research, available online.

[2] (2014). RITA TransStat Origin and Destination Survey database, available online.

[3] Adar, E.; Adamic, L. A. (2005). Tracking information epidemics in blogspace. In: *Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence*. IEEE Computer Society, pp. 207–214.

[4] Aguirre, J.; Papo, D.; Buldú, J. M. (2013). Successful strategies for competing networks. *Nature Physics* **9(4)**, 230–234.

[5] Anderson, P. W.; *et al.* (1972). More is different. *Science* **177(4047)**, 393–396.

[6] Anderson, R. M.; May, R. M.; Anderson, B. (1992). *Infectious diseases of humans: dynamics and control*, vol. 28. Wiley Online Library.

[7] Arthur, W. B. (1999). Complexity and the economy. *science* **284(5411)**, 107–109.

[8] Aurell, E.; Pfitzner, R. (2009). Gaussian belief with dynamic data and in dynamic network. *EPL (Europhysics Letters)* **87(6)**, 68004.

[9] Barabási, A.-L.; Frangos, J. (2002). *Linked : the new science of networks*. Cambridge, Mass: Perseus Pub.

[10] Barrat, A.; Fernandez, B.; Lin, K. K.; Young, L.-S. (2013). Modeling Temporal Networks Using Random Itineraries. *Phys. Rev. Lett.* **110**, 158702.

[11] Battiston, S.; Puliga, M.; Kaushik, R.; Tasca, P.; Caldarelli, G. (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific reports* **2**.

[12] Bavelas, A. (1948). A mathematical model for group structures. *Human organization* **7(3)**, 16–30.

[13] Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the acoustical society of America* .

[14] Berge, C.; Minieka, E. (1973). *Graphs and hypergraphs*, vol. 7. North-Holland publishing company Amsterdam.

[15] Berman, K. A. (1996). Vulnerability of scheduled networks and a generalization of Menger's theorem. *Networks* **28(3)**, 125–134.

[16] Blanchard, P.; Volchenkov, D. (2011). *Random Walks and Difussions on Graphs and Databases*. Springer Berlin Heidelberg.

[17] Blonder, B.; Dornhaus, A. (2011). Time-Ordered Networks Reveal Limitations to Information Flow in Ant Colonies. *PLoS ONE* **6(5)**, e20298.

[18] Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C. I.; Gómez-Gardeñes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports* **544(1)**, 1–122.

[19] Borgatti, S.; Everett, M.; Johnson, J. (2013). *Analyzing Social Networks*. SAGE Publications.

[20] Bornmann, L.; Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* **64(1)**, 45–80.

[21] Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science* **37(1)**, 34–36.

[22] de Bruijn, N. G. (1946). A Combinatorial Problem. *Koninklijke Nederlandse Akademie v. Wetenschappen* **49**, 758–764.

[23] Buldyrev, S. V.; Parshani, R.; Paul, G.; Stanley, H. E.; Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature* **464(7291)**, 1025–1028.

[24] Burgess, R. L. (1969). Communication Networks and Behavioral Consequences! *Human Relations* **22(2)**, 137–159.

[25] Cardillo, A.; Zanin, M.; Gómez-Gardeñes, J.; Romance, M.; del Amo, A. J. G.; Boccaletti, S. (2013). Modeling the multi-layer nature of the European Air Transport Network: Resilience and passengers re-scheduling under random failures. *The European Physical Journal Special Topics* **215(1)**, 23–33.

[26] Checkland, P. (1981). Systems thinking, systems practice .

[27] Chung, F. (2005). Laplacians and the Cheeger Inequality for Directed Graphs. *Annals of Combinatorics* **9**, 1–19.

[28] Cover, T. M.; Thomas, J. A. (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience.

[29] Craven, P.; Wellman, B. (1973). The Network City*. *Sociological inquiry* **43(3-4)**, 57–88.

[30] Csermely, P.; Korcsmáros, T.; Kiss, H. J.; London, G.; Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacology & therapeutics* **138(3)**, 333–408.

[31] D'Agostino, G.; Scala, A. (2014). *Networks of networks: the last frontier of complexity*, vol. 340. Springer.

[32] De Domenico, M.; Solé-Ribalta, A.; Cozzo, E.; Kivelä, M.; Moreno, Y.; Porter, M. A.; Gómez, S.; Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X* **3(4)**, 041022.

[33] De Domenico, M.; Solé-Ribalta, A.; Gómez, S.; Arenas, A. (2014). Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences* **111(23)**, 8351–8356.

[34] Eagle, N.; (Sandy) Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.* **10(4)**, 255–268.

[35] Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics* **69(1)**, 131–152.

[36] Elkana, Y. (1978). Toward a metric of science: The advent of science indicators .

[37] Erdős, P.; Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae (Debrecen)* **6**, 290–297.

[38] Fiedler, M. (1973). Algebraic connectivity of Graphs. *Czechoslovak Mathematical Journal* **23(98)**.

[39] Fiedler, M. (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal* **25(4)**, 619–633.

[40] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry* , 35–41.

[41] Frenken, K.; Ponds, R.; Van Oort, F. (2010). The citation impact of research collaboration in science-based industries: A spatial-institutional analysis. *Papers in regional science* **89(2)**, 351–271.

[42] Gallagher, R.; Appenzeller, T.; Normile, D.; *et al.* (1999). Beyond reductionism. *Science* **284(5411)**, 79.

[43] Gallo, G.; Longo, G.; Pallottino, S.; Nguyen, S. (1993). Directed hypergraphs and applications. *Discrete applied mathematics* **42(2)**, 177–201.

[44] Gallotti, R.; Barthelemy, M. (2015). The multilayer temporal network of public transport in Great Britain. *Scientific data* **2**.

[45] Gao, J.; Buldyrev, S. V.; Havlin, S.; Stanley, H. E. (2011). Robustness of a network of networks. *Physical Review Letters* **107(19)**, 195701.

[46] Gao, J.; Buldyrev, S. V.; Stanley, H. E.; Havlin, S. (2012). Networks formed from interdependent networks. *Nature physics* **8(1)**, 40–48.

[47] Garas, A. (2014). Reaction-Diffusion Processes on Interconnected Scale-Free Networks. *arXiv preprint arXiv:1407.6621* .

[48] Garas, A. (2016). *Interconnected networks*. Springer.

[49] Garas, A.; Garcia, D.; Skowron, M.; Schweitzer, F. (2012). Emotional persistence in online chatting communities. *Scientific Reports* **2**, 402.

[50] Garfield, E.; Merton, R. K. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*, vol. 8. Wiley New York.

[51] Gauffriau, M.; Larsen, P.; Maye, I.; Roulin-Perriard, A.; von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics* **77(1)**, 147–176.

[52] Gfeller, D.; De Los Rios, P. (2007). Spectral coarse graining of complex networks. *Physical review letters* **99(3)**, 038701.

[53] Ghoshal, G.; Zlatić, V.; Caldarelli, G.; Newman, M. (2009). Random hypergraphs and their applications. *Physical Review E* **79(6)**, 066118.

[54] Goldstein, H.; Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* , 385–443.

[55] Gómez, S.; Díaz-Guilera, A.; Gómez-Gardeñes, J.; Pérez-Vicente, C. J.; Moreno, Y.; Arenas, A. (2013). Diffusion Dynamics on Multiplex Networks. *Physical Review Letters* **110(2)**, 028701.

[56] Goodwin, B.; Sole, R. (2000). Signs of Life: How Complexity Pervades Biology.

[57] Grabow, C.; Grosskinsky, S.; Timme, M. (2012). Small-world network spectra in mean-field theory. *Physical review letters* **108(21)**, 218701.

[58] Grinstein, G.; Linsker, R. (2008). Power-law and exponential tails in a stochastic priority-based model queue. *Phys. Rev. E* **77**, 012101.

[59] Gross, T.; D'Lima, C. J. D.; Blasius, B. (2006). Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.* **96**, 208701.

[60] Gross, T.; Sayama, H. (2009). Adaptive Networks. In: T. Gross; H. Sayama (eds.), *Adaptive Networks*, Understanding Complex Systems, Springer Berlin Heidelberg. pp. 1–8.

[61] Harary, F.; Norman, R. (1960). Some properties of line digraphs. *Rendiconti del Circolo Matematico di Palermo* **9(2)**, 161–168.

[62] Hazelkorn, E. (2007). The impact of league tables and ranking systems on higher education decision making. *Higher education management and policy* **19(2)**, 1–24.

[63] Hicks, D.; Wouters, P.; Waltman, L.; de Rijcke, S.; Rafols, I. (2015). Bibliometrics: The Leiden manifesto for research metrics. *Nature* **520**, 429–431.

[64] Hines, P.; Blumsack, S. (2008). A centrality measure for electrical networks. In: *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*. IEEE, pp. 185–185.

[65] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America* **102(46)**, 16569–16572.

[66] Hoffmann, T.; Porter, M. A.; Lambiotte, R. (2013). Random Walks on Stochastic Temporal Networks. In: P. Holme; J. Saramäki (eds.), *Temporal Networks*, Understanding Complex Systems, Springer Berlin Heidelberg. pp. 295–313.

[67] Holme, P. (2003). Network dynamics of ongoing social relationships. *EPL (Europhysics Letters)* **64(3)**, 427.

[68] Holme, P. (2015). Modern temporal network theory: a colloquium. *The European Physical Journal B* **88(9)**, 1–30.

[69] Holme, P.; Saramäki, J. (2012). Temporal networks. *Physics Reports* **519(3)**, 97–125.

[70] Iribarren, J. L.; Moro, E. (2009). Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Phys. Rev. Lett.* **103**, 038702.

[71] Jackson, M. O.; *et al.* (2008). *Social and economic networks*, vol. 3. Princeton university press Princeton.

[72] Jo, H.-H.; Karsai, M.; Kertész, J.; Kaski, K. (2012). Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics* **14(1)**, 013055.

[73] Jo, H.-H.; Perotti, J. I.; Kaski, K.; Kertész, J. (2014). Analytically Solvable Model of Spreading Dynamics with Non-Poissonian Processes. *Phys. Rev. X* **4**, 011041.

[74] Johnson, N. (2009). *Simply Complexity: A clear guide to complexity theory.* Oneworld Publications.

[75] Johnson, S. (2002). *Emergence: The connected lives of ants, brains, cities, and software.* Simon and Schuster.

[76] Kaluza, P.; Kölzsch, A.; Gastner, M. T.; Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society Interface* **7(48)**, 1093–1103.

[77] Karsai, M.; Kivelä, M.; Pan, R. K.; Kaski, K.; Kertész, J.; Barabási, A.-L.; Saramäki, J. (2011). Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 025102.

[78] Karsai, M.; Perra, N.; Vespignani, A. (2014). Time-Varying networks and the weakness of strong ties. *Scientific Reports* , 4001.

[79] Kempe, D.; Kleinberg, J.; Kumar, A. (2000). Connectivity and inference problems for temporal networks. In: *Proceedings of the thirty-second annual ACM symposium on Theory of computing.* ACM, pp. 504–513.

[80] Kempe, D.; Kleinberg, J.; Kumar, A. (2002). Connectivity and Inference Problems for Temporal Networks. *Journal of Computer and System Sciences* **64(4)**, 820 – 842.

[81] Kim, H.; Anderson, R. (2012). Temporal node centrality in complex networks. *Phys. Rev. E* **85(2)**, 1–8.

[82] Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J. P.; Moreno, Y.; Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks* **2(3)**, 203–271.

[83] Klovdahl, A. S. (1985). Social networks and the spread of infectious diseases: the AIDS example. *Social science & medicine* **21(11)**, 1203–1216.

[84] Kostakos, V. (2009). Temporal graphs. *Physica A: Statistical Mechanics and its Applications* **388(6)**, 1007 – 1023.

[85] Kovanen, L.; Karsai, M.; Kaski, K.; Kertész, J.; Saramäki, J. (2011). Temporal motifs in time-dependent networks. *J. Stat. Mech.* **(11)**, P11005.

[86] Lambiotte, R.; Salnikov, V.; Rosvall, M. (2015). Effect of memory on the dynamics of random walks on networks. *Journal of Complex Networks* **3(2)**, 177–188.

[87] Leicht, E.; D'Souza, R. M. (2009). Percolation on interacting networks. *arXiv preprint arXiv:0907.0894* .

[88] Lentz, H. H. K.; Selhorst, T.; Sokolov, I. M. (2013). Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks. *Phys. Rev. Lett.* **110**, 118701.

[89] Leskovec, J.; Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. In: *Proceedings of the 17th international conference on World Wide Web*. ACM, pp. 915–924.

[90] Li, W.; Liu, C.-C.; Zhang, T.; Li, H.; Waterman, M. S.; Zhou, X. J. (2011). Integrative analysis of many weighted co-expression networks using tensor computation. *PLoS Comput Biol* **7(6)**, e1001106.

[91] Little, R. G. (2002). Controlling cascading failure: understanding the vulnerabilities of interconnected infrastructures. *Journal of Urban Technology* **9(1)**, 109–123.

[92] for London, T. (2014). Rolling Origin and Destination Survey (RODS) database.

[93] Lovász, L. (1993). Random walks on graphs: a survey. In: *Combinatorics, Paul Erdös is Eighty (Volume 2), Keszthely (Hungary)*. pp. 1–46.

[94] Lovász, L. (2012). *Large networks and graph limits*, vol. 60. American Mathematical Soc.

[95] MacRoberts, M.; MacRoberts, B. (1996). Problems of citation analysis. *Scientometrics* **36(3)**, 435–444.

[96] MacRoberts, M. H.; MacRoberts, B. R. (1988). Author motivation for not citing influences: A methodological note. *Journal of the American Society for Information Science (1986-1998)* **39(6)**, 432.

[97] Mantegna, R. N.; Stanley, H. E. (1999). *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press.

[98] Martin, T.; Ball, B.; Karrer, B.; Newman, M. (2013). Coauthorship and citation patterns in the Physical Review. *Physical Review E* **88(1)**, 012814.

[99] Martín-Hernández, J.; Wang, H.; Van Mieghem, P.; D'Agostino, G. (2014). Algebraic connectivity of interdependent networks. *Physica A: Statistical Mechanics and its Applications* **404**, 92–105.

[100] Masuda, N.; Klemm, K.; Eguíluz, V. M. (2013). Temporal Networks: Slowing Down Diffusion by Long Lasting Interactions. *Phys. Rev. Lett.* **111**, 188701.

[101] Mayer-Schönberger, V.; Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

[102] Michalski, R.; Palus, S.; Kazienko, P. (2011). Matching organizational structure and social network extracted from email communication. In: *Business Information Systems*. Springer, pp. 197–206.

[103] Milojević, S. (2012). How are academic age, productivity and collaboration related to citing behavior of researchers? *PLoS ONE* **7(11)**, e49176.

[104] Mohar, B.; Alavi, Y.; Chartrand, G.; Oellermann, O. (1991). The Laplacian spectrum of graphs. *Graph theory, combinatorics, and applications* **2**, 871–898.

[105] Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations. .

[106] Morris, M.; Kretzschmar, M. (1995). Concurrent partnerships and transmission dynamics in networks. *Social Networks* **17(3–4)**, 299 – 318.

[107] Mutz, R.; Daniel, H.-D. (2015). What is behind the curtain of the Leiden Ranking? *Journal of the Association for Information Science and Technology* .

[108] Nagatani, T. (2002). The physics of traffic jams. *Reports on progress in physics* **65(9)**, 1331.

[109] Nelson, R. (1995). Probability, stochastic processes, and queueing theory .

[110] Newman, M. (2009). The first-mover advantage in scientific publication. *EPL (Europhysics Letters)* **86(6)**, 68001.

[111] Newman, M. (2010). *Networks: an introduction*. OUP Oxford.

[112] Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical review E* **66(1)**, 016128.

[113] Nicolis, G.; Prigogine, I.; Nocolis, G. (1989). Exploring complexity .

[114] Nicosia, V.; Tang, J.; Mascolo, C.; Musolesi, M.; Russo, G.; Latora, V. (2013). Graph metrics for temporal networks. In: *Temporal Networks*, Springer. pp. 15–40.

[115] Nicosia, V.; Tang, J.; Musolesi, M.; Russo, G.; Mascolo, C.; Latora, V. (2012). Components in time-varying graphs. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **22(2)**, 023101.

[116] Noh, J. D.; Rieger, H. (2004). Random Walks on Complex Networks. *Phys. Rev. Lett.* **92**, 118701.

[117] Nowak, M. A. (2006). *Evolutionary dynamics : exploring the equations of life*. Cambridge, Mass: Belknap Press of Harvard University Press.

[118] Page, L.; Brin, S.; Motwani, R.; Winograd, T. (1999). The PageRank citation ranking: bringing order to the web. .

[119] Pan, R.; Saramäki, J. (2011). Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E* **84(1)**, 1–10.

[120] Pan, R. K.; Kaski, K.; Fortunato, S. (2012). World citation and collaboration networks: uncovering the role of geography in science. *Scientific reports* **2**.

[121] Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

[122] Parshani, R.; Buldyrev, S. V.; Havlin, S. (2010). Interdependent Networks: Reducing the Coupling Strength Leads to a Change from a First to Second Order Percolation Transition. *Physical Review Letters* **105(4)**, 048701.

[123] Parshani, R.; Buldyrev, S. V.; Havlin, S. (2011). Critical effect of dependency groups on the function of networks. *Proceedings of the National Academy of Sciences* **108(3)**, 1007–1010.

[124] Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* **51**, 406–413.

[125] Perra, N.; Baronchelli, A.; Mocanu, D.; Goncalves, B.; Pastor-Satorras, R.; Vespignani, A. (2012). Random Walks and Search in Time-Varying Networks. *Phys. Rev. Lett.* **109**, 238701.

[126] Perra, N.; Gonçalves, B.; Pastor-Satorras, R.; Vespignani, A. (2012). Activity driven modeling of time varying networks. *Scientific reports* **2**.

[127] Pfitzner, R.; Scholtes, I.; Garas, A.; Tessone, C. J.; Schweitzer, F. (2013). Betweenness Preference: Quantifying Correlations in the Topological Dynamics of Temporal Networks. *Phys. Rev. Lett.* **110**, 198701.

[128] Porta, S.; Latora, V.; Wang, F.; Rueda, S.; Strano, E.; Scellato, S.; Cardillo, A.; Belli, E.; Cardenas, F.; Cormenzana, B.; *et al.* (2012). Street centrality and the location of economic activities in Barcelona. *Urban Studies* **49(7)**, 1471–1488.

[129] Porter, M. A.; Gleeson, J. P. (2014). Dynamical Systems on Networks: A Tutorial. *arXiv preprint arXiv:1403.7663* .

[130] Pothen, A.; Simon, H. D.; Liou, K.-P. (1990). Partitioning Sparse Matrices with Eigenvectors of Graphs. *SIAM J. Matrix Anal. Appl.* **11(3)**, 430–452.

[131] Radicchi, F.; Arenas, A. (2013). Abrupt transition in the structural formation of interconnected networks. *Nature Physics* **9(10)**, 1–4.

[132] Rauhvargers, A. (2011). {Global university rankings and their impact}. *Leadership for WorldClass Universities Challenges for Developing Countries* **(June)**.

[133] Ribeiro, B.; Perra, N.; Baronchelli, A. (2013). Quantifying the effect of temporal resolution on time-varying networks. *Scientific reports* **3**.

[134] Rocha, L. E. C.; Blondel, V. D. (2013). Bursts of Vertex Activation and Epidemics in Evolving Networks. *PLoS Comput Biol* **9(3)**, e1002974.

[135] Rocha, L. E. C.; Blondel, V. D. (2013). Flow motifs reveal limitations of the static framework to represent human interactions. *Phys. Rev. E* **87**, 042814.

[136] Rocha, L. E. C.; Liljeros, F.; Holme, P. (2011). Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLoS Comp. Biol.* **7(3)**, e1001109.

[137] Rosato, V.; Issacharoff, L.; Tiriticco, F.; Meloni, S.; Porcellinis, S. D.; Setola, R. (2008). Modelling interdependent infrastructures using interacting dynamical models. *Int. J. of Critical Infrastructures* **4**, 63–79.

[138] Rosenthal, J. S. (1995). Convergence Rates for Markov Chains. *SIAM Review* **37(3)**, pp. 387–405.

[139] Rosvall, M.; Esquivel, A. V.; Lancichinetti, A.; West, J. D.; Lambiotte, R. (2013). Networks with Memory. *ArXiv e-prints* .

[140] Rosvall, M.; Esquivel, A. V.; Lancichinetti, A.; West, J. D.; Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications* **5**.

[141] Salathé, M.; Jones, J. H. (2010). Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput Biol* **6(4)**, e1000736.

[142] Sander, L. M. (2009). *Advanced condensed matter physics*. Cambridge University Press.

[143] Sarigöl, E.; Pfitzner, R.; Scholtes, I.; Garas, A.; Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science* **3(1)**, 1–16.

[144] Scholtes, I.; Wider, N.; Pfitzner, R.; Garas, A.; Tessone, C. J.; Schweitzer, F. (2014). Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nature Communications* **5**, 5024.

[145] Scholtes, Ingo; Wider, Nicolas; Garas, Antonios (2016). Higher-order aggregate networks in the analysis of temporal networks: path structures and centralities*. *Eur. Phys. J. B* **89(3)**, 61.

[146] Sethna, J. P. (2006). Entropy, order parameters, and complexity. *Statistical Mechanics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY* , 14853–2501.

[147] Smith Sidney, L. (1950). Communication pattern and the adaptability of task-oriented groups: an experimental study. *Cambridge, MA: Group Networks Laboratory, Research Laboratory of Electronics, Massachusetts Institute of Technology* .

[148] Sole-Ribalta, A.; De Domenico, M.; Kouvaris, N. E.; Diaz-Guilera, A.; Gomez, S.; Arenas, A. (2013). Spectral properties of the Laplacian of multiplex networks. *Physical Review E* **88(3)**, 032807.

[149] Son, S.-W.; Bizhani, G.; Christensen, C.; Grassberger, P.; Paczuski, M. (2012). Percolation theory on interdependent networks based on epidemic spreading. *EPL (Europhysics Letters)* **97(1)**, 16006.

[150] Sornette, D. (2009). *Why stock markets crash: critical events in complex financial systems*. Princeton University Press.

[151] Starnini, M.; Baronchelli, A.; Barrat, A.; Pastor-Satorras, R. (2012). Random walks on temporal networks. *Phys. Rev. E* **85**, 056115.

[152] Sun, K.; Baronchelli, A.; Perra, N. (2014). Epidemic Spreading in Non-Markovian Time-Varying Networks. *arXiv preprint arXiv:1404.1006* .

[153] Takaguchi, T.; Masuda, N.; Holme, P. (2013). Bursty Communication Patterns Facilitate Spreading in a Threshold-Based Epidemic Dynamics. *PLoS ONE* **8(7)**, e68629.

[154] Takaguchi, T.; Yano, Y.; Yoshida, Y. (2015). Coverage centralities for temporal networks. *ArXiv e-prints* .

[155] Tang, J.; Mascolo, C.; Musolesi, M.; Latora, V. (2011). Exploiting temporal complex network metrics in mobile malware containment. In: *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*. IEEE, pp. 1–9.

[156] Tang, J.; Musolesi, M.; Mascolo, C.; Latora, V. (2010). Characterising temporal distance and reachability in mobile and online social networks. *ACM SIGCOMM Computer Communication Review* **40(1)**, 118–124.

[157] Tang, J.; Musolesi, M.; Mascolo, C.; Latora, V.; Nicosia, V. (2010). Analysing information flows and key mediators through temporal centrality metrics. In: *Proceedings of the 3rd Workshop on Social Network Systems*. ACM, p. 3.

[158] Tang, J.; Scellato, S.; Musolesi, M.; Mascolo, C.; Latora, V. (2010). Small-world behavior in time-varying graphs. *Physical Review E* **81(5)**, 055101.

[159] Tessone, C. J.; Zanette, D. H. (2012). Synchronised firing induced by network dynamics in excitable systems. *EPL* **99(6)**, 68006.

[160] Tononi, G.; Edelman, G. M. (1998). Consciousness and complexity. *science* **282(5395)**, 1846–1851.

[161] Usher, A.; Savino, M. (2006). A World of Difference: A Global Survey of University League Tables. Canadian Education Report Series. *Online Submission* .

[162] Van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics* **62(1)**, 133–143.

[163] Van Regenmortel, M. H. (2004). Reductionism and complexity in molecular biology. *EMBO reports* **5(11)**, 1016–1020.

[164] Vanhems, P.; Barrat, A.; Cattuto, C.; Pinton, J.-F.; Khanafer, N.; Regis, C.; Kim, B.-A.; Comte, B.; Voirin, N. (2013). Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE* **8**, e73970.

[165] Vazquez, A.; Racz, B.; Lukacs, A.; Barabasi, A.-L. (2007). Impact of non-Poissonian activity patterns on spreading processes. *Physical review letters* **98(15)**, 158702.

[166] Vespignani, A. (2010). Complex networks: The fragility of interdependency. *Nature* **464(7291)**, 984–985.

[167] Vinkler, P. (1998). Comparative investigation of frequency and strength of motives toward referencing. The reference threshold model. *Scientometrics* **43(1)**, 107–127.

[168] Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Elsevier.

[169] Wallace, M. L.; Larivière, V.; Gingras, Y. (2012). A small world of citations? The influence of collaboration networks on citation practices. *PloS one* **7(3)**, e33339.

[170] Waltman, L.; Calero-Medina, C.; Kosten, J.; Noyons, E.; Tijssen, R. J.; Eck, N. J.; Leeuwen, T. N.; Raan, A. F.; Visser, M. S.; Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology* **63(12)**, 2419–2432.

[171] Waltman, L.; van Eck, N. J. (2013). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics* **7(4)**, 833–849.

[172] Wang, Z.; Scaglione, A.; Thomas, R. J. (2010). Electrical centrality measures for electric power grid vulnerability analysis. In: *Decision and Control (CDC), 2010 49th IEEE Conference on*. IEEE, pp. 5792–5797.

[173] Watts, D. J. (2004). *Six degrees: The science of a connected age*. WW Norton & Company.

[174] Whitesides, G. M.; Ismagilov, R. F. (1999). Complexity in chemistry. *science* **284(5411)**, 89–92.

[175] Wider, N.; Garas, A.; Scholtes, I.; Schweitzer, F. (2016). An Ensemble Perspective on Multi-layer Networks , 37–59.

[176] Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.

[177] Wu, C. W. (2005). Algebraic connectivity of directed graphs. *Linear and Multilinear Algebra* **53(3)**.

[178] Xuan, B. B.; Ferreira, A.; Jarry, A. (2003). Computing shortest, fastest, and foremost journeys in dynamic networks. *International Journal of Foundations of Computer Science* **14(02)**, 267–285.

[179] Yağan, O.; Gligor, V. (2012). Analysis of complex contagions in random multiplex networks. *Physical Review E* **86(3)**, 036103.

[180] Zhao, K.; Karsai, M.; Bianconi, G. (2011). Entropy of Dynamical Social Networks. *PLoS ONE* **6(12)**, e28116.

[181] Zhao, K.; Stehlé, J.; Bianconi, G.; Barrat, A. (2011). Social network dynamics of face-to-face interactions. *Physical review E* **83(5)**, 056109.